

UNIVERSIDADE FEDERAL DA PARAÍBA
Centro de Ciências Sociais Aplicadas
Departamento de Economia
Programa de Pós-Graduação em Economia

ALÉSSIO TONY CAVALCANTI DE ALMEIDA

**HÁBITOS PREJUDICIAIS À SAÚDE: DEMANDA E
SEUS EFEITOS NO ATRASO ESCOLAR E NO
MERCADO DE TRABALHO**

João Pessoa, Paraíba
2014

ALÉSSIO TONY CAVALCANTI DE ALMEIDA

HÁBITOS PREJUDICIAIS À SAÚDE: DEMANDA E SEUS EFEITOS NO ATRASO ESCOLAR E NO MERCADO DE TRABALHO

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal da Paraíba como parte dos requisitos necessários à obtenção do título de Doutor em Economia.

Orientador: Dr. Ignácio Tavares de Araújo Júnior

João Pessoa, Paraíba

2014

A447h Almeida, Aléssio Tony Cavalcanti de.

Hábitos prejudiciais à saúde: demanda e seus efeitos no atraso escolar e no mercado de trabalho / Aléssio Cavalcanti de Almeida.--João Pessoa, 2014.

150f. : il.

Orientador: Ignácio Tavares de Araújo Júnior

Tese (Doutorado) - UFPB/CCSA

1. Economia. 2. Fatores de risco - produtos não saudáveis.
3. Demanda. 4. Atraso escolar. 5. Penalização salarial.

UFPB/BC

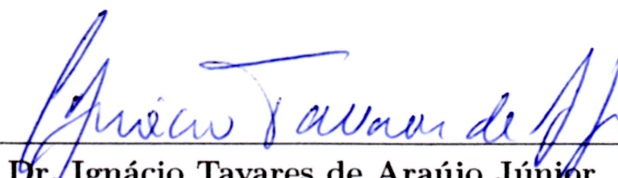
CDU: 33(043)

ALÉSSIO TONY CAVALCANTI DE ALMEIDA

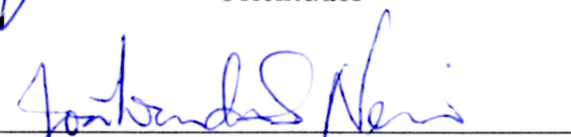
HÁBITOS PREJUDICIAIS À SAÚDE: DEMANDA E SEUS EFEITOS NO ATRASO ESCOLAR E NO MERCADO DE TRABALHO

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Economia da Universidade Federal da Paraíba como parte dos requisitos necessários à obtenção do título de Doutor em Economia. Submetida e APROVADA pela Comissão Examinadora abaixo assinada.

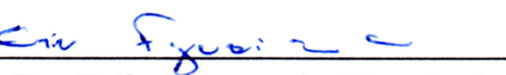
Defesa realizada no Campus I da UFPB em João Pessoa, Paraíba, no dia 27 de Novembro de 2014.




Dr. Ignácio Tavares de Araújo Júnior
Orientador




Dr. José Luis da Silva Netto Júnior
Membro Interno



Dr. Erik Alencar de Figueiredo
Membro Interno



Dr. Francisco de Sousa Ramos
Membro Externo (UFPE)



Dr. Raul da Mota Silveira Neto
Membro Externo (UFPE)

A Deus, minha esposa, meus familiares, amigos e professores que me ajudaram direta e indiretamente a concluir mais esta etapa da minha vida.

Agradecimentos

Um desafio tão grande quanto ao de desenvolver esta tese é o de fazer os agradecimentos, pois a conclusão desta etapa acadêmica reúne não apenas o aprendizado dos últimos anos do curso de doutorado, mas um abrangente processo contínuo de transformação pessoal, acadêmica e profissional.

Inicialmente, agradeço a minha esposa, Ila, pela compreensão e apoio diário, que foram fundamentais para a energia e inspiração necessária para o desenvolvimento desta pesquisa. Não poderia deixar de fazer também um merecido agradecimento aos meus pais, Álvaro e Maria José, pelo suporte durante toda a minha vida, aos meus irmãos e demais familiares – inclusive para aqueles que não habitam mais entre nós – pelas palavras de incentivo.

Ao professor Ignácio Tavares pela oportunidade de transmitir seu conhecimento de pesquisa e o árduo trabalho de orientação, sobretudo no momento crítico de consolidação e estruturação do estudo. Faço um agradecimento também a todos os meus professores que vão além dos ensinamentos dos conteúdos programáticos, como Ivan Targino (DE-UFPB), Alfredo Gomes (IFPB), Carlos Gasparini (MPOG), Nelson Rosas (DE-UFPB), Hilton Martins (DE-UFPB), António Portugal (FEUC), Luiz Renato (DE-UFPB), Erik Figueiredo (DE-UFPB), Sinézio Maia (DE-UFPB), Maria da Conceição (DE-UFPB), José Luis (DE-UFPB), Paulo Amilton (DE-UFPB), Paulo Fernando (DE-UFPB), entre tantos outros que não caberiam neste espaço.

Aos professores Francisco de Sousa Ramos, Raul da Mota Silveira Neto, José Luis da Silva Netto Júnior e Erik Alencar de Figueiredo pela disponibilidade de participar da comissão examinadora e pelas valiosas contribuições para a finalização desta tese.

A todos os colegas do curso de Economia, desde os da época da graduação, pelas experiências e conhecimentos compartilhados. Aos funcionários do PPGE, em especial Risomar Farias, pelo profissionalismo e constante ajuda em cada estágio do curso.

Destaco também a importância do Departamento de Economia da UFPB, pois permitiu-me direcionar exclusivamente os meus esforços para a conclusão do Doutorado. Eu pretendo transformar meus agradecimentos em trabalho para a instituição.

Enfim, um conjunto ainda mais amplo de pessoas contribuíram direta e indiretamente para a realização desta tese, as quais tenho uma profunda gratidão.

Resumo

Esta tese é composta por três ensaios relacionados à demanda por produtos não saudáveis e ao papel de fatores comportamentais de risco à saúde nos resultados escolares e no mercado de trabalho. O [primeiro ensaio](#) analisa a demanda das famílias brasileiras por bebidas alcoólicas e cigarros, com ênfase nas elasticidades preço e dispêndio e nas simulações de mudanças nos preços destes itens sobre o bem-estar. A abordagem *Quadratic Almost Ideal Demand System* e os dados da Pesquisa de Orçamentos Familiares 2008-2009 e do Suplemento de Tabagismo da Pesquisa Nacional por Amostra de Domicílios 2008 são usados neste estudo. Pelos resultados auferidos, cigarro e bebidas alcoólicas possuem elasticidade-dispêndio positiva e relação de substitutibilidade independentemente do nível de renda domiciliar *per capita* e da macrorregião de residência. Choques positivos nos preços desses itens possuem um baixo ajustamento de demanda, assim como a taxa de compensação de renda requerida em função de mudanças nos preços do cigarro é maior para as famílias e regiões mais ricas. Por sua vez, o [segundo ensaio](#) objetiva avaliar o papel da exposição de alunos aos fatores de risco comportamentais à saúde – cigarro, bebida alcoólica e excesso de peso – no resultado educacional no Brasil. Para tanto, os microdados da Pesquisa Nacional de Saúde do Escolar 2012 e técnicas paramétricas e não-paramétricas são utilizados para a estimação do efeito da exposição a esses fatores sobre o indicador de atraso escolar de discentes no 9º ano do ensino fundamental. Os principais resultados da pesquisa confirmam a hipótese de que a exposição aos fatores de risco se relaciona de forma direta com o atraso escolar, sendo os efeitos mais intensos para os estudantes com pior nível socioeconômico. Esses achados ratificam a importância de políticas públicas que promovam a prevenção desses fatores de risco desde a infância, pois as consequências dessas exposições geram desdobramentos não apenas na saúde mas também no componente educacional do capital humano. Por fim, o propósito central do [último ensaio](#) é explorar a heterogeneidade da repercussão de hábitos pessoais não saudáveis, expressa pelo uso do cigarro, sobre a produtividade do trabalho e no *trade-off* salário-risco. Com base na Pesquisa Especial de Tabagismo integrante da Pesquisa Nacional por Amostra de Domicílios 2008 e no Anuário Estatístico de Acidentes do Trabalho 2008, os modelos empíricos são desenvolvidos por meio de regressores quantílicos com variáveis instrumentais. Os resultados encontrados explicitam que a penalização salarial do cigarro com controle para endogeneidade é estatisticamente significativa ao longo da distribuição dos rendimentos individuais, com perdas salariais variando de 15,2% a 36,5%. Os fumantes também recebem um menor prêmio ao risco no comparativo ao recebido pelos não-usuários do cigarro nas atividades econômicas com maior incidência de acidentes do trabalho não-fatais. Conforme essas estimativas, o valor estatístico de um acidente do trabalho por ano é, na mediana, para os não-fumantes de aproximadamente R\$ 6,4 mil por acidente e R\$ 3,5 mil para os fumantes, com diferenças também nesses valores nos outros quantis da distribuição condicional do salário.

Palavras-chave: Fatores de risco. Demanda. Atraso escolar. Penalização salarial.

Abstract

This thesis consists of three essays related to demand for unhealthy products and the role of behavioral risk factors to health on school outcomes and the labor market. The [first essay](#) analyzes the demand of Brazilian families for alcoholic beverages and cigarettes, with emphasis on price and expenditure elasticities and simulations of changes in the prices of these items on the welfare. We use the Quadratic Almost Ideal Demand System and data from the Household Budget Survey 2008-2009 and the Smoking Supplement of the National Research by Household Sample 2008. The main results show that cigarettes and alcohol have positive expenditure elasticity of demand and substitution relationship in terms of cross-price, regardless of per capita income level and region of residence. Positive price changes in these items have low adjustment of demand, as well as the rate of required income compensation due to changes in cigarette prices is higher for richer households and regions. In turn, the [second essay](#) evaluates the role of exposure of students to behavioral risk factors to health – smoking, alcohol and overweight – in educational attainment in Brazil. We use microdata from the National Survey of School Health 2012 provided by Brazilian Institute of Geography and Statistics and parametric and nonparametric techniques to estimate the effect of exposure to these factors in the indicator of delay in school progression of students in the 9th grade of elementary school. The main results confirm the hypothesis that exposure to risk factors has direct effect on delay in school progression. Furthermore, these effects are more intense for students with lower socioeconomic level. Then, the findings of this study ratify the importance of public policies that promote prevention of these risk factors among children, once the exposure to risk factors to health generates repercussions not only in health but also in the educational component of human capital. Finally, the core purpose of the [last essay](#) is to explore the heterogeneity of the repercussion of unhealthy personal behaviors, expressed by cigarette smoking, on labor productivity and wage-risk trade-off. Based on the Special Smoking Survey included in the National Survey by Household Sample 2008 and Yearbook Statistics of Job Injuries 2008, the empirical models are developed by instrumental quantile regression. The findings show that the smoking wage penalty with endogeneity control is statistically significant over the distribution of labor income, with wage losses ranging from 15.2% to 36.5%. Furthermore, smokers receive a lower risk premium than nonsmokers in economic activities with higher incidence of nonfatal occupational injuries. According to these estimates, the value of a statistical injury per year is, on median, for non-smokers approximately R\$ 6,400 per injury and R\$ 3,500 for smokers, with differences also in the other quantiles of the conditional wage distribution.

Keywords: Risk factors. Demand. School delay. Wage penalty.

Lista de ilustrações

Figura 2.1 –Taxa de incidência anual de neoplasias malignas no pulmão, traqueia e brônquios por 100.000 habitantes em 2008-2009, segundo gênero e regiões do Brasil	21
Figura 2.2 –Parcela gasta, em média, com bebidas alcoólicas e cigarro por decil da renda domiciliar <i>per capita</i> nas macrorregiões brasileiras	38
Figura 2.3 –Elasticidades estimadas para os modelos sem e com ajustamento para consumo censurado e endogeneidade das despesas totais, bebidas e cigarro no Brasil 2008-2009	43
Figura 2.4 –Elasticidades-preço das demandas Marshallianas para bebidas alcoólicas e cigarro (ambos sentidos) por decil da renda domiciliar <i>per capita</i> . . .	47
Figura 2.5 –Variação compensatória com ajustamento de demanda para os consumidores de bebidas alcoólicas e cigarro, conforme unidade federativa (Perda de bem-estar) – em %	50
Figura 2.6 –Variação compensatória com ajustamento de demanda no Brasil por famílias pobres, ricas e total ao longo dos meses de aplicação da POF – em %	51
Figura 3.1 –Taxa de distorção idade-série no 9º ano do Ensino Fundamental por dependência administrativa e grandes regiões do Brasil a partir da amostra final deste trabalho	68
Figura 3.2 –Taxa de exposição aos fatores de risco à saúde por parte de estudantes do 9º ano do Ensino Fundamental no Brasil a partir da amostra final deste trabalho	71
Figura 3.3 –Efeito médio dos fatores de risco comportamentais à saúde sobre o atraso escolar – não condicionado ao contrafactual observável	74
Figura 3.4 –Efeito dos fatores de risco comportamentais à saúde sobre o atraso escolar por NSE no Brasil 2012 – não condicionado ao contrafactual observável	77
Figura 3.5 –Efeito dos fatores de risco comportamentais à saúde sobre o atraso escolar por decis do NSE no Brasil 2012 – condicionado ao contrafactual observável	81
Figura 4.1 –Relação salário-risco por status de fumante para dois casos hipotéticos	93
Figura 4.2 –Comparação dos efeitos do tabagismo por quantil da distribuição condicional dos rendimentos do trabalhador no Brasil entre RQ e RQVI, 2008	114

Figura A.1 –Curvas não-paramétricas de Engel por grupos de produtos alimentares, bebidas e cigarro no Brasil 2008-2009	137
Figura A.2 –Resíduos da equação reduzida das despesas totais por decil de renda . .	138
Figura B.1 –Curvas não-paramétricas entre os termos residuais das equações do estágio inicial do PSM	144
Figura B.2 –Densidade Kernel do <i>propensity score</i> entre expostos e não-expostos por tipo de fator de risco comportamental à saúde	145

Lista de tabelas

Tabela 2.1	–Evidências da literatura para as elasticidades dispêndio e preço do cigarro e de bebidas alcoólicas	24
Tabela 2.2	–Estatísticas descritivas das variáveis usadas na pesquisa e a proporção de despesas zero por categoria de produto	35
Tabela 2.3	–Características da amostra por grupos de famílias demandantes ($w_i > 0$) e não-demandantes por cigarro e bebidas alcoólicas no Brasil	36
Tabela 2.4	–Parâmetros do QUAIDS ajustado aos valores censurados e endogeneidade das despesas totais	39
Tabela 2.5	–Elasticidade-dispêndio da demanda por decil de renda domiciliar <i>per capita</i> e por níveis regionais no Brasil 2008-2009	41
Tabela 2.6	–Matriz de elasticidades-preço das demandas Marshallianas e Hicksianas no Brasil 2008-2009	44
Tabela 2.7	–Variação compensatória com e sem efeito substituição (ES) para os consumidores de bebidas alcoólicas e cigarro, conforme decil de renda e macrorregiões (Perda de bem-estar) – em %	49
Tabela 3.1	–Resultados da estimação do modelo logit para a probabilidade de atraso escolar dos estudantes no Brasil 2012	75
Tabela 3.2	–Resultados da estimação dos coeficientes para o <i>propensity score</i> por tipo de exposição ao fator de risco à saúde da criança no Brasil 2012	78
Tabela 3.3	–Efeito dos fatores de risco comportamentais à saúde sobre o atraso escolar por diferentes técnicas de pareamento no Brasil 2012 – condicionado ao contrafactual observável	80
Tabela 3.4	–Análise de sensibilidade aos fatores não-observados por tipo de exposição e diferentes técnicas de pareamento nos estimadores por PSM	83
Tabela 4.1	–Resultados sobre penalização salarial do tabagismo na literatura especializada	89
Tabela 4.2	–Estatísticas descritivas das variáveis utilizadas nos modelos de penalização e diferenciação compensatória	104
Tabela 4.3	–Características da amostra por grupos de fumantes e não-fumantes no Brasil, PETab 2008	106
Tabela 4.4	–Resultados da regressão para a penalização salarial do fumante corrente no Brasil, PETab 2008. Variável dependente: logaritmo do salário-hora	108
Tabela 4.5	–Testes para endogeneidade da variável H_i no MQ2E	111

Tabela 4.6 –Resultados para a penalização salarial do fumante corrente por RQ no Brasil, PETab 2008. Variável dependente: logaritmo do salário-hora . . .	112
Tabela 4.7 –Resultados para a penalização salarial do fumante corrente por RQVI no Brasil, PETab 2008. Variável dependente: logaritmo do salário-hora . . .	113
Tabela 4.8 –Coeficiente do prêmio ao risco de acidentes do trabalho e VSI no Brasil por status de fumante e diferentes modelagens econométricas, 2008 . . .	116
 Tabela A.1 –Estimativas do modelo probit multivariado – Estágio de correção dos valores censurados	138
Tabela A.2 –Matriz de correlação dos termos de erros das equações no modelo probit multivariado – Estágio de correção dos valores censurados	139
Tabela A.3 –Parâmetros do QUAIDS sem ajustes para consumo censurado e endogeneidade das despesas totais	140
 Tabela B.1 –Descrição dos itens utilizados para a elaboração do Nível Socioeconômico (NSE)	141
Tabela B.2 –Descrição dos itens utilizados para a elaboração do indicador de infraestrutura da escola (IEE)	142
Tabela B.3 –Estatística descritivas das variáveis	142
Tabela B.4 –Efeito marginal de variáveis explicativas discretas (D) e contínuas (C) sobre a probabilidade de atraso escolar no Brasil 2012 – a partir do modelo logit do Estágio I	143
 Tabela C.1 –Descrição das estatísticas do total de acidentes do trabalho (AT) no Brasil por atividade econômica e do índice de risco de acidentes do trabalho (IRAT), 2008	146
Tabela C.2 –Resultados para a penalização salarial do fumante corrente para especificações com ocupação informal e ramos da atividade econômica – MQO e MQ2E. Variável dependente: logaritmo do salário-hora	147
Tabela C.4 –Análise de Robustez – diferentes especificações de instrumentos para os modelos de penalização e diferenciação compensatória	148
Tabela C.3 –Definição das variáveis usadas na pesquisa por tipo de modelo	149
Tabela C.3 –Definição das variáveis usadas na pesquisa por tipo de modelo	150

Lista de abreviaturas e siglas

AEAT	Anuário Estatístico de Acidentes do Trabalho
AIDS	Almost Ideal Demand System
BLS	Bureau of Labor Statistics
CAT	Comunicações de Acidentes do Trabalho
CIA	Hipótese de independência condicional
COFINS	Contribuição para o Financiamento da Seguridade Social
CNAE	Classificação Nacional de Atividades Econômicas
CR	Com reposição
DCNT	Doenças Crônicas Não-Transmissíveis
DIS	Distorção idade-série
EAP	Expected a posteriori
ENEM	Exame Nacional do Ensino Médio
EPM	Efeito parcial médio
EPPM	Efeito parcial na média
ERIC	EconLit e Education Resources Information Center
FDA	Função de Distribuição Acumulada
FDP	Função Densidade de Probabilidade
FGLNS	Feasible Generalized Nonlinear Least Squares
IBGE	Instituto Brasileiro de Geografia e Estatística
IEE	Indicador de infraestrutura escolar
IGP-DI	Índice Geral de Preços - disponibilidade interna
IMC	Índice de Massa Corporal
INCA	Instituto Nacional de Câncer

IPCA	Índice de Preço ao Consumidor Amplo
IPi	Imposto sobre Produtos Industrializados
IRAT	Índice de risco de acidente do trabalho
MPS	Ministério da Previdência Social
MQ2E	Mínimos Quadrados em dois estágios
MQO	Mínimos Quadrados Ordinários
MS	Ministério da Saúde
NELS	National Education Longitudinal Study
NLSAH	National Longitudinal Study of Adolescent Health
NLSUR	Nonlinear Seemingly Unrelated Regression
NLSY	National Longitudinal Survey of Youth
NSE	Nível socioeconômico
OMS	Organização Mundial da Saúde
PASEP	Programa de Formação do Patrimônio do Servidor Público
PENSE	Pesquisa Nacional de Saúde do Escolar
PETab	Pesquisa Especial de Tabagismo
PIGLOG	Price Independent Generalized Logarithmic
PIS	Programa de Integração Social
PNAD	Pesquisa Nacional por Amostra de Domicílios
POF	Pesquisa de Orçamentos Familiares
PPV	Pesquisa sobre Padrões de Vida
PSM	Propensity Score Matching
QUAIDS	Quadratic Almost Ideal Demand System
RQ	Regressão Quantílica
RQVI	Regressão Quantílica com Variáveis Instrumentais
SR	Sem reposição

SUS	Sistema Único de Saúde
TOA	Total de ocupados em todas as outras atividades na mesma divisão econômica, exceto a atividade indivíduo i
TRI	Teoria de Resposta ao Item
UPA	Unidade primária de amostragem
USA	Unidade secundária de amostragem
USAFA	U.S. Air Force Academy
UTA	Unidade terciária de amostragem
VC	Variação Compensatória
VI	Variáveis Instrumentais
VSI	Value of a Statistical Injury
VSL	Value of a Statistical Life
WHO	World Health Organization

Sumário

1	Introdução	17
2	Demanda por bebidas alcoólicas e cigarros no Brasil: elasticidades, microsimulação e variações no bem-estar	20
2.1	Introdução	20
2.2	Sistema de demanda	22
2.2.1	Quadratic Almost Ideal Demand System (QUAIDS)	25
2.2.2	QUAIDS com ajustamento sociodemográfico	26
2.3	Estratégia empírica	27
2.3.1	Consumo censurado e endogeneidade das despesas	28
2.3.2	Estimação da equação das parcelas de gastos e das elasticidades	30
2.3.3	Variação compensatória	31
2.4	Base de dados e descrição das variáveis	32
2.5	Resultados	36
2.5.1	Característica das famílias demandantes por cigarro e bebidas alcoólicas e estimativas do modelo QUAIDS ajustado	36
2.5.2	Elasticidades	40
2.5.3	Mudanças nos preços e variação do bem-estar	48
2.6	Conclusões	52
3	Efeitos da exposição aos fatores de risco comportamentais à saúde sobre o atraso escolar no Brasil	54
3.1	Introdução	54
3.2	Função de produção educacional e o papel das condições de saúde	55
3.3	Estratégia empírica	60
3.3.1	Modelo logit: efeito médio não condicionado ao contrafactual observável	62
3.3.2	Modelo PSM: efeito médio condicionado ao grupo de controle observável	63
3.4	Base de dados	66
3.4.1	Distorção idade-série	67
3.4.2	Indicadores de exposição aos fatores de risco comportamentais	69
3.4.3	Demais variáveis	70
3.5	Resultados	72
3.5.1	Exploração inicial: Determinantes da atraso escolar	73
3.5.2	Efeito médio da exposição usando PSM	76

3.5.3	Análise de sensibilidade: qual o tamanho do viés de fatores não-observados?	82
3.6	Conclusões	84
4	Tabagismo, penalização salarial e diferenciação compensatória para risco de acidentes no mercado de trabalho brasileiro	86
4.1	Introdução	86
4.2	Revisão da literatura sobre tabagismo e retornos no mercado de trabalho	88
4.2.1	Diferenciação compensatória salarial para trabalhadores fumantes	92
4.3	Estratégia empírica	95
4.3.1	Modelo econométrico para penalização salarial	95
4.3.2	Modelo econométrico: diferenciação compensatória salarial por status de fumante	99
4.4	Base de dados e descrição das variáveis	102
4.5	Resultados	105
4.5.1	Penalização salarial para usuários do cigarro	105
4.5.1.1	Resultados por MQO e MQ2E	107
4.5.1.2	Resultados por Regressão quantílica e RQVI	111
4.5.2	Diferenciação compensatória salarial para atividades perigosas	115
4.6	Conclusões	119
5	Considerações Finais	121
	Referências	123
A	Apêndice – Primeiro Ensaio	137
B	Apêndice – Segundo Ensaio	141
C	Apêndice – Terceiro Ensaio	146

1 Introdução

A World Health Organization (WHO, 2010) ressalta que a condição de saúde individual é determinada por uma série de canais. Por um lado, ela é reflexo das circunstâncias materiais – como habitação, potencial de consumo e ambiente de trabalho –, herança genética e qualidade do sistema de serviços de saúde. Por outro, o elevado número de mortes atribuídas às Doenças Crônicas Não Transmissíveis (DCNT) fazem das preferências individuais um fator crucial para a formação do estoque do capital saúde presente e futuro da população, bem como um problema de saúde pública. Em 2008, as DCNT, lideradas pelos fatores de risco modificáveis – tabagismo, consumo de bebidas alcoólicas, sedentarismo, baixa ingestão de frutas e vegetais – foram responsáveis por 63% dos óbitos no mundo e por 80% nos países em desenvolvimento (WHO, 2011).

Além dos custos irreparáveis (a vida), a adoção de hábitos não saudáveis, que segundo Grossman (1972) reduz o investimento líquido do estoque de saúde, possui impacto na dimensão econômica individual e social. Nessa direção, existe uma preocupação crescente em se compreender os determinantes dessas escolhas (JIMENEZ; LABEAGA, 1994; GOEL; MOREY, 1995; EAKINS; GALLAGHER, 2003; CUTLER; GLAESER, 2005; KOKSAL; WOHLGENANT, 2013), inclusive meditando a regra desempenhada pelas posturas prejudiciais à saúde ao longo do ciclo de vida do agente, tanto em termos dos resultados na escola (JAMISON, 1986; GLEWWE; JACOBY, 1995; JEYNES, 2002; CROSNOW; MULLER, 2004; LINDO; SWENSEN; WADDELL, 2013) quanto dos resultados no mercado de trabalho (HERSCH; VISCUSI, 1990; LEVINE; GUSTAFSON; VELENCHIK, 1997; VISCUSI; HERSCH, 2001; BOCKERMAN; HYYTINEN; KAPRIO, 2014).

Muito embora o Brasil possua um quantitativo considerável de crianças expostas aos fatores comportamentais de risco à saúde e uma elevada parcela da população em idade produtiva enquadrada como obesa e usuária do cigarro, ainda são escassos trabalhos que integram as literaturas de economia da saúde (GROSSMAN, 1972) e vício racional¹ (BECKER; MURPHY, 1988; CHALOUPIKA, 1991) com análises de demanda, educação e mercado de trabalho. Segundo a Pesquisa de Orçamentos Familiares (POF) 2008-2009, 17% dos indivíduos com 25 anos de idade ou mais são obesos², enquanto que mais de 1/3 da população entre 25 e 60 anos de idade é atual ou ex-usuária do cigarro, conforme a Pesquisa Especial de Tabagismo (PETab) integrante da Pesquisa Nacional por Amostra de Domicílios (PNAD) 2008. Quanto às crianças, a Pesquisa Nacional de Saúde do Escolar

¹ Para Becker & Murphy (1988), essa teoria admite que as decisões de consumo dos agentes por produtos de vício (como cigarro, álcool, heroína etc.) podem ser modeladas, pois suas prospectivas de maximização ocorrem em um arcabouço de preferências estáveis (previsíveis).

² A definição de obesidade utilizada é dada pelo Índice de Massa Corporal (IMC) igual ou maior que $30\text{kg}/\text{m}^2$.

(PENSE) 2012 reporta, por exemplo, que 1/5 e mais de 3/5 dos estudantes da última fase do ensino fundamental já tiveram experiência, respectivamente, com cigarro e com bebidas alcoólicas.

Nessa conjuntura, existe uma lacuna de trabalhos no Brasil que analisem a relação entre os comportamentos prejudiciais à saúde dos indivíduos e os seus desdobramentos em diferentes fases da vida (na infância e na fase produtiva), bem como investiguem a demanda específica por produtos atrelados aos fatores de risco à saúde. Assim, esta tese desenvolve três ensaios, organizados em capítulos, que contemplam essa agenda de pesquisa, alicerçada nos fundamentos da economia da saúde e nos custos implícitos das preferências não saudáveis.

No primeiro ensaio, com base nos trabalhos de [Banks, Blundell & Lewbel \(1997\)](#), [Shonkwiler & Yen \(1999\)](#), [Blundell & Robin \(1999\)](#), [Friedman & Levinsohn \(2002\)](#), [Yen \(2005\)](#), [Zheng & Henneberry \(2010\)](#), [Wood, Nelson & Nogueira \(2012\)](#) e [Bilgic & Yen \(2013\)](#), o escopo do estudo é conduzido pela avaliação da demanda por bebidas alcoólicas e cigarro, produtos vinculados aos fatores de risco modificáveis, através de um sistema de demanda do tipo *Quadratic Almost Ideal Demand System* (QUAIDS) ajustado para características sociodemográficas, consumo censurado e endogeneidade das despesas totais. Os parâmetros das equações de demanda, que subsidiam os cálculos das elasticidades dispendio e preço, são desenvolvidos através de um modelo de sistemas de equações não-lineares. A partir de um exercício de microsimulação em decorrência de uma mudança nos preços das bebidas alcoólicas e cigarro, os valores da variação compensatória de Hicks são também obtidos para o caso brasileiro.

O segundo ensaio evidencia como a exposição aos fatores de risco comportamentais à saúde na juventude, captado por uso de cigarro, consumo de bebidas alcoólicas e excesso de peso, influenciam os resultados dos alunos no ambiente da escola. Esse estágio de pesquisa tem como principais referências [Glewwe & Jacoby \(1995\)](#), [Yamada, Kendix & Yamada \(1996\)](#), [Ding *et al.* \(2009\)](#), [Eide, Showalter & Goldhaber \(2010\)](#), [Cawley & Ruhm \(2011\)](#), [Ponzo \(2013\)](#) e [Lindo, Swensen & Waddell \(2013\)](#), que associam a relação entre as condições de saúde na infância e a performance acadêmica. Este trabalho pretende avançar nas discussões do papel dos hábitos não saudáveis dos alunos, na fase inicial dos estudos, sobre os resultados educacionais com uma modelagem não-paramétrica, recortes amostrais específicos e análise de sensibilidade dos coeficientes aos fatores não-observados.

Finalmente, o último ensaio objetiva examinar o papel de preferências de risco à saúde, capturada pelo tabagismo, por parte de agentes na fase produtiva sobre os resultados no mercado de trabalho, no que se referem à penalização salarial e aos diferenciais compensatórios para atividades arriscadas entre fumantes e não-fumantes. Trabalhos como [Hersch & Viscusi \(1990\)](#), [Levine, Gustafson & Velenchik \(1997\)](#), [Viscusi & Hersch \(2001\)](#), [Munasinghe & Sicherman \(2006\)](#), [Grafova & Stafford \(2009\)](#), [Anger & Kvasnicka \(2010\)](#) e

Bockerman, Hyytinen & Kaprio (2014) norteiam o desenvolvimento desta terceira parte da pesquisa. Além do mais, a heterogeneidade do custo indireto do tabagismo e do prêmio ao risco de acidentes do trabalho é adicionada na análise por meio da abordagem de regressores quantílicos com variáveis instrumentais.

Além desta introdução, esta tese é composta por mais quatro capítulos. O [Capítulo 2](#), que se trata do primeiro ensaio, aborda a demanda por bebidas alcoólicas e cigarro. O [Capítulo 3](#) mostra os efeitos da exposição ao cigarro, bebidas alcoólicas e excesso de peso dos estudantes no 9º ano do ensino fundamental sobre o atraso escolar. O [Capítulo 4](#), equivalente ao último ensaio, apresenta os custos monetários indiretos do tabagismo e evidências sobre o *trade-off* salário-risco entre fumantes e não-fumantes no mercado de trabalho brasileiro. Por fim, o [Capítulo 5](#) traz as considerações finais.

2 Demanda por bebidas alcoólicas e cigarros no Brasil: elasticidades, microssimulação e variações no bem-estar

2.1 Introdução

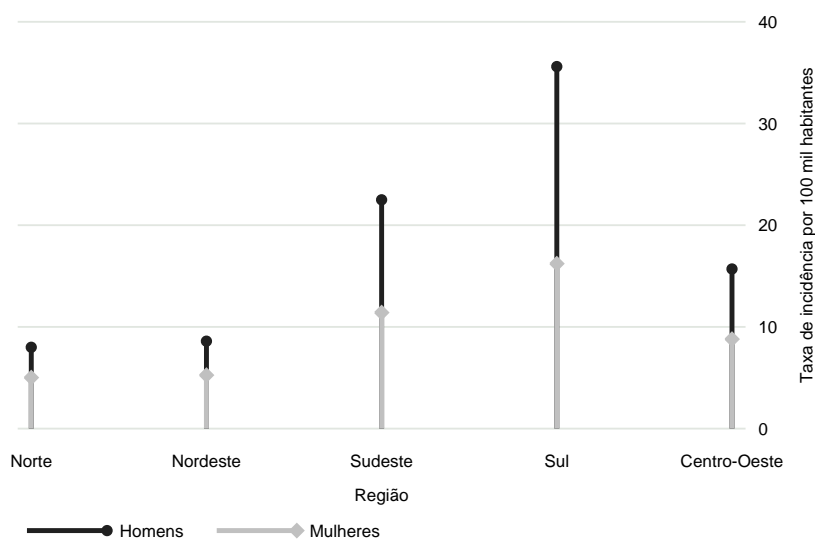
As literaturas, teórica e empírica, enfatizam a necessidade de se compreender a demanda dos consumidores por produtos não saudáveis, como o cigarro e bebidas alcoólicas, por eles gerarem desdobramentos diretos na própria saúde e de terceiros e uma série de externalidades negativas com fortes implicações sociais e econômicas (BECKER; MURPHY, 1988; SAFFER; CHALOUKKA, 1994; CHALOUKKA; WARNER, 1999). A partir da utilização de um modelo de demanda, os estudos estimam a sensibilidade desses produtos em respostas às variações nos seus preços e na renda, usando dados agregados, como em Barten (1964), Johnson & Oksanen (1977), Saffer & Chaloupka (1994), Carvalho & Lobao (1998), e desagregados, como em Jimenez & Labeaga (1994), Duffy (1995), Goel & Morey (1995), Decker & Schwartz (2000), Gil & Molina (2009), Yamamoto (2011), Carvalho, Siqueira & Nogueira (2013). No tocante aos estudos sobre demanda por cigarro e bebidas alcoólicas no Brasil, ainda são escassas pesquisas que trabalhem com estes produtos de forma conjunta em um sistema de demanda, considerando aspectos regionais, estratificação de renda e repercussão de choques nos preços desses itens sobre as famílias demandantes³.

É importante destacar que o Brasil possui diferentes padrões regionais de consumo para bebidas com teor alcoólico e cigarro. Segundo dados da Pesquisa de Orçamentos Familiares (POF) 2008-2009, as regiões Sul e Sudeste apresentam mais de 22% de domicílios com despesas em bebidas alcoólicas e acima de 25% para o cigarro. No Nordeste, a proporção de domicílios que gastam com alcoólicos e cigarro, respectivamente, é 16,4% e 19,4%, sendo que o peso no orçamento desses itens para as famílias nordestinas, sobretudo para o cigarro, é bem abaixo da média para o eixo Centro-Sul. Como um dos possíveis resultados disso, segundo as estatísticas sobre a taxa de incidência anual de neoplasias malignas no pulmão, traqueia e brônquios para 2008-2009 dispostas na Figura 2.1, as regiões com maior consumo de cigarro registram também as maiores taxas de neoplasias malignas por 100 mil habitantes tanto para as ocorrências entre homens quanto entre mulheres.

Compreender a demanda dos indivíduos por produtos não saudáveis, inclusive

³ Para o caso brasileiro, Carvalho & Lobao (1998), Menezes, Silveira & Azzoni (2008), Pintos-Payeras (2009) e Yamamoto (2011) são exemplos de trabalhos que direta ou indiretamente analisam a demanda por bebidas alcoólicas e/ou cigarro no país.

Figura 2.1 – Taxa de incidência anual de neoplasias malignas no pulmão, traqueia e brônquios por 100.000 habitantes em 2008-2009, segundo gênero e regiões do Brasil



Fonte: Elaboração própria a partir das estimativas desenvolvidas pelo Ministério da Saúde/Instituto Nacional do Câncer (INCA).

verificando a sensibilidade desta demanda às mudanças nos preços e na renda, é relevante para o desenho de políticas públicas que visem desestimular a procura por eles. No Brasil, políticas de cunho tarifário, que ampliam os impostos de bebidas com teor alcoólico e cigarro, e não-tarifário, como a lei nº 9.294/96 que proíbe o consumo do cigarro em ambientes fechados e a lei nº 11.705/08 – conhecida como Lei Seca –, vêm dando uma atenção especial aos chamados fatores de risco modificáveis a saúde. Nesse contexto, o presente estudo analisa a demanda por bebidas alcoólicas e cigarro, considerando um vetor de consumo que inclui outros itens alimentares. Com base nessa abordagem, constrói-se uma matriz de elasticidade-preço da demanda compensada e não-compensada, assim como a elasticidade-dispêndio da demanda para diferentes estratos de renda domiciliar *per capita* e região de localização. Não obstante, objetiva-se também identificar os impactos de modificações nos preços do cigarro e do álcool sobre o bem-estar das famílias por meio da medida monetária de variação compensatória de Hicks, com e sem inclusão do efeito substituição, como realizado por [Friedman & Levinsohn \(2002\)](#), [Vu & Glewwe \(2011\)](#) e [Wood, Nelson & Nogueira \(2012\)](#).

Para o cumprimento dos objetivos acima, estima-se um sistema de demanda do tipo *Quadratic Almost Ideal Demand System* (QUAIDS), que abrange preços, renda, características sociodemográficas e ajustes para endogeneidade da despesa total e consumo censurado. Na estimação desse sistema, os microdados da POF 2008-2009 são utilizados em conjunto com o suplemento especial de tabagismo da Pesquisa Nacional por Amostra de Domicílios (PNAD) 2008.

Este capítulo está dividido em seis partes, o que inclui esta introdução. A [Seção 2.2](#) descreve o sistema de demanda, incluindo a abordagem QUAIDS com ajustamento sociodemográfico. A [Seção 2.3](#) e a [Seção 2.4](#) detalham o modelo empírico e a base de dados da pesquisa. Ao final, a [Seção 2.5](#) e a [Seção 2.6](#) apresentam os principais resultados e as considerações finais.

2.2 Sistema de demanda

A análise de demanda modela o padrão de gastos das famílias sobre um dado grupo de bens e serviços com o objetivo de estimar elasticidades preços e renda, aferir o bem-estar dos consumidores e subsidiar a elaboração de políticas públicas e investimentos privados na economia (POI, 2002). Esse tipo de exame pode ser desenvolvido através de um arcabouço microeconômico baseado em um sistema de equações de demanda, tais como os sistemas: de despesas lineares, de funções de demanda de Rotterdam, de utilidade logarítmica transcendental (translog), *Almost Ideal Demand System* (AIDS) e do AIDS em sua versão quadrática, apresentados, respectivamente, em [Stone \(1954\)](#), [Barten \(1964, 1969\)](#), [Christensen, Jorgenson & Lau \(1975\)](#), [Deaton & Muellbauer \(1980\)](#) e [Banks, Blundell & Lewbel \(1997\)](#).

O modelo proposto por [Stone \(1954\)](#) visa estabelecer um sistema de equações de demanda que sejam factíveis de implementação, meditando fundamentos teóricos e estatísticos. Com base em dados anuais para o Reino Unido entre 1920 e 1938 e para seis grupos de produtos (carnes e produtos diários, frutas e verduras, bebidas e tabaco, despesas domiciliares – como aluguel, energia, combustível etc. – e todos os outros consumos de bens e serviço), o estudo mostra a importância da análise empírica do padrão de consumo para fins de elaboração de políticas e previsão de consumo. Em igual direção, [Barten \(1964\)](#) propõe reduzir a lacuna entre a teoria de demanda do consumidor e a pesquisa empírica, a partir da utilização de um modelo teórico que impõe restrições microeconômicas aos parâmetros estimados pela análise de regressão, usando séries temporais de despesas dos consumidores e preços para 14 bens e serviços para a Holanda entre os períodos de 1921-1939 e 1948-1958.

O ponto de partida para estudos econômicos sobre a demanda do consumidor, como ressalta [Christensen, Jorgenson & Lau \(1975\)](#), é um sistema que envolve a quantidade consumida de cada produto em função das despesas totais e dos preços de todas as *commodities*. As funções de demanda devem ser consistentes com a maximização da utilidade, em que os parâmetros do modelo econométrico precisam obedecer às condições previstas pela teoria. Esses autores incluem também em suas análises a hipótese de homoteticidade, verificando mudanças nos parâmetros estimados para despesas de consumo pessoal de bens duráveis e não-duráveis e outros serviços nos EUA de 1929 a 1972.

Desde o desenvolvimento do modelo AIDS por [Deaton & Muellbauer \(1980\)](#) e o avanço dos recursos computacionais no período, uma grande variedade de estudos empíricos, baseados nesse sistema de equações de demanda, avaliam o comportamento do consumidor para diferentes grupos de bens, com registro em uma diversidade de países. [Deaton & Muellbauer \(1980\)](#) aplicam o modelo AIDS para despesas agregadas dos consumidores em oito grupo de bens não-duráveis (alimentos, roupas, moradia, combustíveis, bebidas e tabaco, transportes e comunicações, outros bens e serviços), com dados britânicos anuais de 1954 a 1974, identificando a elasticidade renda e preços destas *commodities*.

[Blundell, Pashardes & Weber \(1993\)](#) avaliam a estrutura dos sistemas de demandas microeconômicas e o papel dos atributos do agregado familiar no comportamento do consumidor. Os microdados da pesquisa *British Family Expenditure Survey* de 1970 a 1984 e uma extensão quadrática do modelo AIDS são utilizados, com os dados dispostos em cortes transversais agrupados contendo aproximadamente 4 mil domicílios ao longo de cada um dos 15 anos. A variabilidade nos padrões de demanda observada em dados desagregados, com alta heterogeneidade de atributos familiares e diferentes níveis de renda, fazem com que os autores incluam na especificação do sistema de demanda a não-linearidade do padrão das despesas dos consumidores. Desta forma, o trabalho de [Banks, Blundell & Lewbel \(1997\)](#) formaliza o modelo AIDS com a extensão quadrática, conhecido na literatura como QUAIDS.

Os trabalhos mais recentes que usam o sistema de demanda, como [Shonkwiler & Yen \(1999\)](#), [Yen \(2005\)](#), [Schlindwein & Kassouf \(2006\)](#), [Kebede \(2008\)](#), [Yen, Yuan & Liu \(2009\)](#), [Zheng & Henneberry \(2010\)](#), [Tefera, Demeke & Rashid \(2012\)](#), [Carvalho, Siqueira & Nogueira \(2013\)](#), [Bilgic & Yen \(2013\)](#), [Oliver \(2014\)](#), chamam atenção para a necessidade de corrigir as equações de demanda para problemas relativos ao número excessivo de despesas nulas, devido à aplicação das pesquisas de orçamentos que ocorrem dentro de um intervalo bastante limitado de tempo para cada família. Para essa literatura, tal fato pode resultar em um problema de seletividade amostral, tornando os parâmetros das equações das composições dos gastos potencialmente viesados.

Outro problema empírico dessa área é o de endogeneidade das despesas totais, especialmente quando se trabalha com apenas uma parte do vetor de consumo ([BLUNDELL; ROBIN, 1999](#); [ZHENG; HENNEBERRY, 2010](#)). A teoria pressupõe que a renda familiar é exógena aos dispêndios, porém na prática para assegurar que $\sum_i w_i = 1$, onde w_i é a parcela gasta com o produto i em relação aos gastos totais, assume-se que a renda é estabelecida pelo total de despesas familiares ao longo do número restrito de bens avaliados ([BLUNDELL; ROBIN, 1999](#)). Com efeito, a despesa total (m) é potencialmente endógena no modelo QUAIDS, ou seja, a variável explanatória m pode ser correlacionada com o termo de erro (ϵ) da especificação da demanda, $Cov(m, \epsilon) \neq 0$, principalmente para os casos em que o vetor de consumo está desagregado.

Quando se trata especificamente dos estudos sobre a demanda para itens como bebidas alcoólicas e cigarro, existem pesquisas que objetivam de forma indireta, como Barten (1964), Deaton & Muellbauer (1980), Fan, Cramer & Wailes (1994), Menezes, Silveira & Azzoni (2008) e Pintos-Payeras (2009), e direta, como Jimenez & Labeaga (1994), Duffy (1995), Carvalho & Lobao (1998), Yen (2005), Gil & Molina (2009) e Yamamoto (2011), desenvolver equações de demanda para os supracitados produtos, de modo a obter informações sobre elasticidade-preço e dispêndio. A Tabela 2.1 sumariza alguns resultados internacionais e nacionais acerca das elasticidades-preço (do próprio produto) e dispêndio relativos à tais produtos.

Tabela 2.1 – Evidências da literatura para as elasticidades dispêndio e preço do cigarro e de bebidas alcoólicas

Estudo	Modelo	Período	Local	Dados	Dispêndio		Preço	
					Cigarro	Álcool	Cigarro	Álcool
Barten (1964)	Rotterdam	1921-1939 1948-1958	Holanda	S	0,50	1,50	-0,53	-0,44
Johnson & Oksanen (1977)	Linear	1956-1971	Canadá	P	–	0,002 ^a	–	-0,27
Blundell, Pashardes & Weber (1993)	QUAIDS	1970-1984	UK	CP	–	2,29	–	-1,58
Jimenez & Labeaga (1994)	AIDS	1980-1981	Espanha	C	0,14	–	-0,48	–
Fan, Cramer & Wailes (1994)	AIDS-D	1957-1990	China	S	1,73	3,39	0,22	-0,80
Duffy (1995)	Rotterdam	1963-1992	UK	S	0,96	0,884 ^a	-0,351	-0,03
	AIDS	1963-1992	UK	S	1,05	1,07 ^a	-0,42	-0,24
	AIDS-D	1963-1992	UK	S	1,01	1,44 ^a	-0,47	-0,29
Wang <i>et al.</i> (1996)	Sintético	1987-1988	EUA	C	–	1,10 ^a	–	-0,52
Carvalho & Lobao (1998)	U-míope	1983-1994	Brasil	S	0,31 ^b	–	-0,20	–
	U-racional	1983-1994	Brasil	S	0,23 ^b	–	-0,14	–
Duffy (2003)	AIDS-D	1963-1996	UK	S	0,40	0,98 ^a	0,41	-0,41
Eakins & Gallagher (2003)	AIDS	1960-1988	Irlanda	S		0,77 ^a		-0,42
	AIDS-D*	1960-1988	Irlanda	S		0,16 ^a		-0,53
	AIDS	1987-1989	Brasil	P	0,30	–	-2,84	–
Menezes <i>et al.</i> (2008)		1995-1996						
Kebede (2008)	QUAIDS	1994-1995 1997	Etiopia	P	–	1,24	–	-0,54
Pintos-Payeras (2009)	AIDS	2002-2003	Brasil	C	0,63	1,38	-1,03	-1,11
Gil & Molina (2009)	QUAIDS	2000	Espanha	C	–	1,02 ^a	–	-0,57
Yamamoto (2011)	AIDS	2008-2009	Brasil	C	–	1,11 ^a	–	-0,21

Fonte: Elaboração própria.

Legenda: AIDS-D = modelo AIDS dinâmico (com resultados de curto-prazo, quando for o caso); U = modelo univariado (com apenas uma equação); S = análise agregada em nível de país em séries de tempo; C = corte transversal; CP = corte transversal em *pooling*; P = dados em painel; *a* = Referente ao consumo de cerveja; *b* = elasticidade-renda e curto-prazo.

A Tabela 2.1, de forma geral, mostra que o consumo do cigarro tem uma baixa elasticidade quanto às mudanças nos preços e no gasto total quando comparado aos valores da elasticidade para a demanda por bebidas. No caso do Brasil, são raros os trabalhos que tratam de forma conjunta a demanda por esses produtos. Carvalho & Lobao (1998) e Menezes, Silveira & Azzoni (2008) encontram que elasticidade-dispêndio do cigarro no país é de aproximadamente 0,30, enquanto que para Pintos-Payeras (2009) este valor é de 0,63. Em todos os casos, alterações na renda dos agentes provocam um baixo efeito na demanda por cigarro. Já em termos de mudanças nos preços, os achados empíricos desses estudos são divergentes, particularmente quando se observa a elasticidade preço

de Menezes, Silveira & Azzoni (2008). Para bebidas alcoólicas, Pintos-Payeras (2009) e Yamamoto (2011) obtêm, respeitada esta ordem, as seguintes estimativas: um aumento na renda em 10% causa uma elevação no consumo em 13,8% e 11,1%; já 10% a mais nos preços, reduz a demanda em 11,1% e 2,1%. Dos estudos para o Brasil na Tabela 2.1, Pintos-Payeras (2009) é o único que traz informações sobre as elasticidades preço e dispêndio para os dois grupos de produtos, utilizando a POF 2002-2003 com uma amostra restringida para algumas regiões metropolitanas do país.

Na finalidade de complementar a literatura, em especial, sobre a demanda por bebidas alcoólicas e cigarro no Brasil, este trabalho utiliza o modelo QUAIDS como base para o processo de estimação das elasticidades e de medidas de variação de bem-estar resultantes de mudanças nos preços dos citados produtos. A escolha desta abordagem é motivada por dois pontos centrais: atendimento dos pressupostos básicos exigidos pela teoria do consumidor; não-linearidade da curva de Engel⁴, porquanto análises mais desagregadas da demanda apontam uma maior heterogeneidade no comportamento das despesas familiares (BLUNDELL; PASHARDES; WEBER, 1993). As próximas duas subseções descrevem as principais características do modelo QUAIDS.

2.2.1 Quadratic Almost Ideal Demand System (QUAIDS)

No modelo QUAIDS, a função utilidade indireta dos consumidores para n produtos, gasto total m e vetor de preços \mathbf{p} , que tem base nas preferências da classe *Price Independent Generalized Logarithmic*⁵ (PIGLOG), são apresentadas na Equação 2.1

$$\ln V(\mathbf{p}, m) = \left\{ \left[\frac{\ln m - \ln a(\mathbf{p})}{b(\mathbf{p})} \right]^{-1} + \lambda(\mathbf{p}) \right\}^{-1}, \quad (2.1)$$

onde: m é a despesa total da família; $a(\mathbf{p})$, $b(\mathbf{p})$ e $\lambda(\mathbf{p})$ são funções do vetor de preços \mathbf{p} ; $\left[\frac{\ln m - \ln a(\mathbf{p})}{b(\mathbf{p})} \right]$ representa a função utilidade indireta do sistema de demanda PIGLOG. Dada a propriedade de homogeneidade da função utilidade indireta, tem-se que $a(\mathbf{p})$ é homogênea de grau um em p e as outras duas funções, $b(\mathbf{p})$ e $\lambda(\mathbf{p})$, homogêneas de grau zero para os preços. Mais precisamente, tem-se que: $b(\mathbf{p})$ é um agregador de preços tipo Cobb-Douglas dado por $b(\mathbf{p}) = \prod_{i=1}^n p_i^{\beta_i}$, com $\sum_i \beta_i = 0$; $\lambda(\mathbf{p})$ é expresso por $\lambda(\mathbf{p}) = \sum_{i=1}^n \lambda_i \ln p_i$, com

⁴ Pela Figura A.1 no Apêndice, que ilustra o formato da curva de Engel para nove grupos de produtos alimentares, bebidas e cigarro para o caso brasileiro, tem-se uma sinalização de não-linearidade entre a composição dos itens na cesta de consumo e o nível de despesa das famílias. A significância do parâmetro λ associado ao termo quadrático na seção de resultados confirmará a importância da inclusão do termo quadrático da despesa total no sistema de demanda.

⁵ As preferências tipo PIGLOG são definidas por: $\ln e(\mathbf{p}, u) = \ln a(\mathbf{p}) + u \ln b(\mathbf{p})$, onde \mathbf{p} é o vetor de preços, $e(\cdot)$ representa a função dispêndio necessária para atingir o nível de utilidade u e as variáveis $a(\mathbf{p})$ e $b(\mathbf{p})$ captam os custos para o alcance do nível de utilidade (MUELLBAUER, 1976).

$\sum_i \lambda_i = 0$; e o $\ln a(\mathbf{p})$ é definido na Equação 2.2 por uma função agregadora de preço logarítmica transcendental (translog),

$$\ln a(\mathbf{p}) = \alpha_0 + \sum_j \alpha_j \ln p_j + \frac{1}{2} \sum_i \sum_j \gamma_{ij} \ln p_i \ln p_j \quad (2.2)$$

sendo, $\sum_j \alpha_j = 1$ e $\sum_i \sum_j \gamma_{ij} = 0$.

Ao aplicar a identidade de Roy para a função utilidade indireta, a Equação 2.3 reproduz a composição dos gastos de uma família com o i -ésimo produto w_i no modelo QUAIDS,

$$w_i = \alpha_i + \sum_{j=1}^n \gamma_{ji} \ln(p_j) + \beta_i \ln \left[\frac{m}{a(\mathbf{p})} \right] + \frac{\lambda_i}{b(\mathbf{p})} \left\{ \ln \left[\frac{m}{a(\mathbf{p})} \right] \right\}^2. \quad (2.3)$$

Pela Equação 2.3, a parcela dos gastos da h -ésima família com o produto i é uma função dos preços do próprio item e de todos os outros, incluindo a renda (representado pela despesa total ajustada pelos índices de preços) em nível e ao quadrado. Os parâmetros β e λ capturam o efeito da renda no consumo dos produtos e γ indica a relação da demanda pelo produto com o próprio preço e os preços dos outros bens, no qual eles fornecem as bases para o cálculo das elasticidades. Admitindo as condições de aditividade, homogeneidade e simetria impostas pela teoria (DEATON; MUELLBAUER, 1980; BANKS; BLUNDELL; LEWBEL, 1997), esses parâmetros são usados para o cálculo das elasticidades da demanda em relação aos preços e ao dispêndio.

2.2.2 QUAIDS com ajustamento sociodemográfico

Estudos como Pollak & Wales (1981), Ray (1983), Blundell, Pashardes & Weber (1993) e Blacklow, Nicholas & Ray (2010) ressaltam a influência dos fatores contextuais, demográficos e sociais no comportamento do consumidor, cujo acréscimo do vetor de variáveis sociodemográficas no sistema é relevante para não gerar resultados viesados. Pollak & Wales (1981) apontam que a composição etária da família e o seu tamanho são importantes elementos nos determinantes do padrão de gastos, pois famílias com mais crianças tendem a apresentar um comportamento distinto de famílias com uma maior composição de adultos.

A Equação 2.3, com base na descrição de Poi (2012), pode ser reescrita usando o método de Ray (1983) para ajustar a equação de composição dos gastos do modelo QUAIDS às características sociodemográficas. Supõe-se que a função dispêndio com ajuste ao vetor de características sociodemográficas (\mathbf{Z}) é dada pela relação entre a função dispêndio de referência $e^R(\mathbf{p}, u)$ e um fator de escala das características sociodemográficas $[\bar{m}_0(\mathbf{Z}) \times \phi(\mathbf{p}, \mathbf{Z}, u)]$, ou seja: $e(\mathbf{p}, \mathbf{Z}, u) = [\bar{m}_0(\mathbf{Z}) \times \phi(\mathbf{p}, \mathbf{Z}, u)] \times e^R(\mathbf{p}, \mathbf{Z}, u)$. O primeiro termo do fator de escala diz respeito a mudanças na função dispêndio da família como uma

resposta do vetor \mathbf{Z} , independentemente de mudanças no padrão de consumo, enquanto o segundo termo capta mudanças nos preços relativos e no padrão de consumo corrente⁶. Baseado na incorporação desse fator de escala na função dispêndio, a Equação 2.3 é atualizada para a expressão a seguir:

$$w_i = \alpha_i + \sum_{j=1}^n \gamma_{ji} \ln(p_j) + (\beta_i + \boldsymbol{\eta}'_i \mathbf{Z}) \ln \left[\frac{m}{\bar{m}_0(\mathbf{Z})a(\mathbf{p})} \right] + \frac{\lambda_i}{b(\mathbf{p})c(\mathbf{p}, \mathbf{Z})} \left\{ \ln \left[\frac{m}{\bar{m}_0(\mathbf{Z})a(\mathbf{p})} \right] \right\}^2, \quad (2.4)$$

onde: $\bar{m}_0(\mathbf{Z}) = 1 + \boldsymbol{\rho}'\mathbf{Z}$ representa os efeitos das características sociodemográficas sobre a despesa total, sendo $\boldsymbol{\rho}$ o vetor a ser estimado; $\boldsymbol{\eta}$ é o ajuste para as mudanças relativas nas despesas com cada item i de produtos, com $\sum_{i=1}^n \eta_i = 0$; $c(\mathbf{p}, \mathbf{Z}) = \prod_{i=1}^n p_i^{\eta'_i \mathbf{Z}}$ é um agregador de preços com ponderação do vetor de variáveis sociodemográficas.

De posse dos parâmetros advindos do sistema apresentado, Banks, Blundell & Lewbel (1997) sugerem que as interpretações sejam feitas em cima das elasticidades, requerendo inicialmente que a Equação 2.4 seja derivada em termos do logaritmo da despesa total ($\ln m$) e do logaritmo do preços ($\ln p_j$), como pode ser observado nas Equações 2.5 e 2.6.

$$\mu_i \equiv \frac{\partial w_i}{\partial \ln m} = \beta_i + \boldsymbol{\eta}'_i \mathbf{Z} + \frac{2\lambda_i}{b(\mathbf{p})c(\mathbf{p}, \mathbf{Z})} \left\{ \ln \left[\frac{m}{\bar{m}_0(\mathbf{Z})a(\mathbf{p})} \right] \right\} \quad (2.5)$$

$$\mu_{ij} \equiv \frac{\partial w_i}{\partial \ln p_j} = \gamma_{ij} - \mu_i \left(\alpha_j + \sum_k \gamma_{jk} \ln p_k \right) - \frac{\lambda_i(\beta_j + \boldsymbol{\eta}'_j \mathbf{Z})}{b(\mathbf{p})c(\mathbf{p}, \mathbf{Z})} \left\{ \ln \left[\frac{m}{\bar{m}_0(\mathbf{Z})a(\mathbf{p})} \right] \right\}^2 \quad (2.6)$$

A partir da Equação 2.5, a elasticidade-dispêndio em relação ao bem i é dada por $e_i = \mu_i/w_i + 1$. A Equação 2.6 é usada para a obtenção das elasticidades-preços Marshallianas (ou não-compensadas), que são definidas por $e_{ij}^u = \mu_{ij}/w_i - \delta_{ij}$, em que δ_{ij} é o delta de Kronecker, que admite $\delta_{ij} = 1$, se $i = j$, e $\delta_{ij} = 0$, caso contrário. As elasticidades-preços Hicksianas (ou compensadas) podem ser calculadas utilizando a equação de Slutsky, por: $e_{ij}^c = e_{ij}^u + w_j e_i$.

2.3 Estratégia empírica

Como o presente estudo analisa a demanda das famílias com ênfase no consumo de cigarro e bebidas alcoólicas, o modelo QUAIDS exibido na Equação 2.4, incorporando o

⁶ Segundo Poi (2012), essa função pode ser parametrizada pela seguinte expressão: $\ln \phi(\mathbf{p}, \mathbf{Z}, u) = \frac{\prod_j p_j^{\beta_j} \left(\prod_j p_j^{\eta'_j \mathbf{Z}} - 1 \right)}{\frac{1}{u} - \sum_j \lambda_j \ln p_j}$.

vetor de variáveis sociodemográficas, serve como referência principal para os resultados desta pesquisa. Contudo, a literatura chama atenção para dois problemas empíricos na estimação do modelo QUAIDS: número expressivo de observações com despesas nulas nas pesquisas de orçamento (HEIEN; WESSELLS, 1990; SHONKWILER; YEN, 1999); endogeneidade das despesas totais (BLUNDELL; ROBIN, 1999).

Para contornar essas questões, a estimação dos parâmetros do sistema de demanda é feita em três estágios. No primeiro estágio, os valores censurados para as despesas são tratados através do procedimento de Shonkwiler & Yen (1999). A segunda etapa corrige o problema de endogeneidade das despesas totais, baseando-se em Blundell & Robin (1999), que incorporam os resíduos estimados na equação das despesas totais no sistema de demanda. Ao final dessas duas etapas, estima-se o modelo QUAIDS ajustado aos fatores sociodemográficos, variáveis censuradas e endogeneidade das despesas, para enfim analisar as elasticidades e mudanças no bem-estar para os produtos relacionados com o consumo de dois fatores de risco modificáveis à saúde: cigarro e bebidas alcoólicas.

As duas próximas subseções fornecem mais detalhes das etapas de correção, bem como a subseção 2.3.3 apresenta os procedimentos adotados para a mensuração do impacto de mudanças nos preços do cigarro e/ou bebidas alcoólicas no bem-estar dos consumidores.

2.3.1 Consumo censurado e endogeneidade das despesas

Para os valores censurados das despesas em diferentes categorias de produtos, é possível encontrar duas estratégias para correção deste viés na amostra em um sistema de demanda: aplicação do procedimento similar ao de Heckman (1979), no qual a razão inversa de Mills obtido por um modelo probit univariado é incluído como regressor nas equações de demanda, como um instrumento que incorpora os traços latentes da variável censurada (HEIEN; WESSELLS, 1990); uso do procedimento de Shonkwiler & Yen (1999), que também se baseia na utilização de um modelo de probabilidade no primeiro estágio com a diferença na especificação do ajuste nas equações de demanda. No caso desta pesquisa, adota-se a abordagem de Shonkwiler & Yen (1999), com a utilização de um modelo probit multivariado, obtido através do método de máxima verossimilhança, que possui um melhor ajuste para contabilizar possíveis correlações entre as diferentes categorias de produtos (ZHENG; HENNEBERRY, 2010). Assim, o processo de correção é realizado da seguinte forma:

- Estimação de um modelo de probabilidade simultaneamente para todas as *commodities* tendo como variável dependente o indicador de consumo da família para cada grupo de produto (d_{ih}), com $d_{ih} = 1$, se a família apresenta composição de despesa positiva ($w_{ih} > 0$) na i -ésima categoria de bens, e $d_{ih} = 0$, caso contrário. Supondo

um nível de utilidade não-observável para a decisão de consumo (d_{hi}^*), a especificação do modelo é apresentada na Equação 2.7

$$d_{hi}^* = \boldsymbol{\pi}_i' \mathbf{J}_{hi} + \kappa_{hi}, \text{ sendo } d_{hi} = 1 \text{ se } d_{hi}^* > 0; d_{hi} = 0 \text{ c.c.} \quad (2.7)$$

$$\begin{pmatrix} \kappa_{h1} \\ \kappa_{h2} \\ \vdots \\ \kappa_{hn} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & r_{12} & \dots & r_{1n} \\ r_{21} & 1 & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \dots & 1 \end{pmatrix} \right],$$

onde: \mathbf{J} é o vetor de variáveis explicativas formada pela despesa total ($\ln m$), pelo vetor de preços ($\ln \mathbf{p}$), pelas características sociodemográficas \mathbf{Z} e *dummies* de sazonalidade (carnaval, festas juninas e natal) que podem afetar a decisão da família pelo consumo do i -ésimo produto no período, assim como variáveis relativas ao status de ocupação da pessoa de referência e a taxa de moradores ocupados no domicílio que também se relacionam com tais decisões; $\boldsymbol{\pi}$ é o vetor de parâmetros. Essa expressão representa um modelo probit multivariado, em que κ_{hi} são os termos de erros padronizados que assumem uma forma funcional normal multivariada. O coeficiente de correlação que capta a relação entre os termos de erros estocásticos das equações é indicado por r , no qual a sua significância estatística ($r \neq 0$) explicita que as decisões de consumo não são feitas independentemente.

- Com base nas estimativas da Equação 2.7, calcula-se a função de distribuição acumulada (FDA), $\Phi_{hi} = \Phi(\boldsymbol{\pi}_i' \mathbf{J}_{hi})$, e a função de densidade de probabilidade (FDP), $\varphi_{hi} = \varphi(\boldsymbol{\pi}_i' \mathbf{J}_{hi})$, para cada categoria de produto e para cada família, que são incluídas na especificação ampliada do QUAIDS. O parâmetro τ_i representa a covariância entre o termo de erro no modelo QUAIDS e o termo de erro do probit multivariado para as decisões de consumo. Dado o valor observado da parcela gasta W_{ih} e o valor latente desse dispêndio w_{ih} determinado pela Equação 2.4, o modelo QUAIDS ajustado para o consumo censurado é escrito como:

$$W_{hi} = \Phi(\hat{\boldsymbol{\pi}}_i' \mathbf{J}_{hi}) \times \hat{w}_{hi} + \hat{\tau}_i \varphi(\hat{\boldsymbol{\pi}}_i' \mathbf{J}_{hi}), \text{ com } W_{hi} \geq 0. \quad (2.8)$$

Calculadas a FDA e a FDP para correção descrita anteriormente, o próximo passo é tratar da endogeneidade das despesas totais da família. Com base em [Blundell & Robin \(1999\)](#) e [Zheng & Henneberry \(2010\)](#), o problema da endogeneidade das despesas é contornado utilizando os resíduos da Equação 2.9 para as despesas totais como instrumento a ser incluído no sistema de demanda,

$$\ln m_h = a_0 + \boldsymbol{\Lambda}' \mathbf{Z}_h + \boldsymbol{\Lambda}' \ln \mathbf{p}_h + e_y \ln Y_h + v_h, \quad (2.9)$$

onde: \mathbf{Z} e \mathbf{p} são, respectivamente, os vetores de variáveis sociodemográficas e de preços; Y_h representa a renda total domiciliar e e_y a elasticidade renda para a despesa total (m_h).

2.3.2 Estimação da equação das parcelas de gastos e das elasticidades

As equações de demanda com correção para os problemas descritos na subseção anterior têm seus parâmetros calculados através de um sistema de equações não-lineares ajustados pelo método *feasible generalized nonlinear least squares* (FGNLS), através de um processo iterativo – similar ao da estimação por máxima verossimilhança com erros estocásticos normais multivariados⁷.

Destaca-se que as correções implementadas, fazem com que a condição de aditividade não seja plenamente assegurada (YEN; LIN; SMALLWOOD, 2003). Para superar esse problema, os autores propõem que a n -ésima categoria de produtos seja tratada como um grupo residual⁸, ou seja: $W_{hn} = 1 - \sum_{i=1}^{n-1} W_{hi}$. As elasticidades preço e dispêndio para o produto omitido são derivadas das condições do modelo QUAIDS que implicam: (i) $\sum_{i=1}^n W_i e_i = 0$; (ii) $\sum_{i=1}^n W_i e_{ij}^u = -W_j$; (iii) $\sum_{j=1}^N e_{ij}^u + e_i = 0$. A Equação 2.10 define o sistema de demanda objeto de estimação

$$\begin{aligned} W_{h1} &= \Phi_{h1} \left\{ \alpha_1 + \sum_{j=1}^9 \gamma_{1j} \ln(p_{hj}) + (\beta_1 + \boldsymbol{\eta}'_1 \mathbf{Z}_h) \ln \left(\frac{m_h}{\zeta(\mathbf{p}, \mathbf{Z})} \right) + \frac{\lambda_1}{b(\mathbf{p})c(\mathbf{p}, \mathbf{Z})} \left[\ln \left(\frac{m_h}{\zeta(\mathbf{p}, \mathbf{Z})} \right) \right]^2 \right\} + \tau_{h1} \varphi_{h1} + \epsilon_{h1} \\ W_{h2} &= \Phi_{h2} \left\{ \alpha_2 + \sum_{j=1}^9 \gamma_{2j} \ln(p_{hj}) + (\beta_2 + \boldsymbol{\eta}'_2 \mathbf{Z}_h) \ln \left(\frac{m_h}{\zeta(\mathbf{p}, \mathbf{Z})} \right) + \frac{\lambda_2}{b(\mathbf{p})c(\mathbf{p}, \mathbf{Z})} \left[\ln \left(\frac{m_h}{\zeta(\mathbf{p}, \mathbf{Z})} \right) \right]^2 \right\} + \tau_{h2} \varphi_{h2} + \epsilon_{h2} \\ \vdots &= \vdots \\ W_{h8} &= \Phi_{h8} \left\{ \alpha_8 + \sum_{j=1}^9 \gamma_{8j} \ln(p_{hj}) + (\beta_8 + \boldsymbol{\eta}'_8 \mathbf{Z}_h) \ln \left(\frac{m_h}{\zeta(\mathbf{p}, \mathbf{Z})} \right) + \frac{\lambda_8}{b(\mathbf{p})c(\mathbf{p}, \mathbf{Z})} \left[\ln \left(\frac{m_h}{\zeta(\mathbf{p}, \mathbf{Z})} \right) \right]^2 \right\} + \tau_{h8} \varphi_{h8} + \epsilon_{h8}, \end{aligned} \quad (2.10)$$

onde: $\zeta(\mathbf{p}, \mathbf{Z}) = \bar{m}_0(\mathbf{Z}_h) \times a_h(\mathbf{p})$; $\epsilon_{hi} = \vartheta_i v_h + \nu_{hi}$. Como os erros da h -ésima observação, $\nu_{h1}, \nu_{h2}, \dots, \nu_{h(n-1)}$, podem ser correlacionados, as $n-1$ equações são ajustadas em conjunto para tornar as estimativas mais eficientes. Conforme Banks, Blundell & Lewbel (1997), o valor de α_0 , necessário para calibrar a equação agregadora de preço translog é conhecido como o nível mínimo de gasto necessário para subsistência, cujo valor pode ser estabelecido pelo valor mínimo do logaritmo da despesa total das famílias. No caso do corrente estudo o valor usado é 4,43 (\approx R\$84,00).

De posse dos parâmetros estimados na Equação 2.10, calcula-se as elasticidades com os devidos ajustes através das seguintes expressões:

- Elasticidade-dispêndio: $E_i = e_i \times \Phi_i$;

⁷ Nesta pesquisa, o comando *nlsur* do *software* Stata foi usado para obtenção desses parâmetros, com adaptação da *function evaluator program*, provido por Poi (2002) e Poi (2008), para os ajustes no desenho amostral, na despesa censurada e na endogeneidade das despesas totais.

⁸ Neste trabalho, a categoria residual é representada pelo grupo de produtos denominado como outros alimentos.

- Elasticidade-preço da demanda não-compensada: $E_{ij}^u = \frac{\mu_{ij}}{W_i} \times \Phi_i + \varphi_i \times \pi_{ij} (1 - \frac{\tau_i}{W_i}) - \delta_{ij}$. O parâmetro π_{ij} é associado ao preço da j-ésima categoria no primeiro estágio (modelo probit multivariado);
- Elasticidade-preço da demanda compensada, pela equação de Slutsky: $E_{ij}^c = E_{ij}^u + W_j \times E_i$.

2.3.3 Variação compensatória

Esta parte apresenta um indicador que mensura o efeito de alterações nos preços do cigarro e de bebidas alcoólicas sobre o bem-estar. Quando se avalia o efeito de uma mudança na alíquota tributária ou um choque de oferta que provoca mudanças nos preços dos produtos, é pertinente investigar o seu impacto sobre o bem-estar dos consumidores com medidas passíveis de comparação entre os agentes (SMALL; ROSEN, 1981).

Dentre as medidas mais usadas para tais fins, a literatura empírica sublinha a variação compensatória (VC), que capta o valor necessário para reparar as perdas do consumidor após uma variação positiva nos preços, indicando o quanto de renda adicional o consumidor deveria receber para manter o nível de utilidade idêntico ao observado na situação inicial (sem alteração de preços). Teoricamente, a VC requer basicamente o nível de utilidade inicial u^0 e dois vetores de preços: \mathbf{p}^0 , cenário inicial; \mathbf{p}^1 , com ao menos uma mudança de preço diferente de zero. A Equação 2.11 a seguir mostra essa relação:

$$VC_h = e(\mathbf{p}_h^1, u_h^0, \bar{\mathbf{Z}}_h) - e(\mathbf{p}_h^0, u_h^0, \bar{\mathbf{Z}}_h). \quad (2.11)$$

Pela Equação 2.11, o impacto da mudança de preço de um ou mais produtos no bem-estar dos consumidores é identificado supondo constantes as características socio-demográficas. Friedman & Levinsohn (2002), Vu & Glewwe (2011), Tefera, Demeke & Rashid (2012) e Wood, Nelson & Nogueira (2012) aplicam a expansão da série de Taylor no mínimo da função dispêndio em relação ao preço para gerar uma aproximação do valor que compensaria os consumidores pós-mudança de preços, de modo a restaurar o nível de utilidade pré-mudança.

Para a obtenção da VC, Friedman & Levinsohn (2002) sugerem o uso da aproximação de segunda ordem da série de Taylor, pois a aproximação de primeira ordem não pondera os efeitos de ajustamento da demanda, assumindo como nulo o efeito substituição (ES). Dada a expansão de segunda ordem e a aplicação do lema de Shepard em conjunto com a permuta da demanda Hicksiana pela Marshalliana no ponto ótimo, a VC aproximada

é definida por:

$$VC_h \cong \left(\sum_{i=1}^n W_{hi} \frac{\Delta p_{hi}}{p_{hi}^0} + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n E_{hij}^c W_{hi} \frac{\Delta p_{hi}}{p_{hi}^0} \frac{\Delta p_{hj}}{p_{hj}^0} \right) \times m_h^0, \quad (2.12)$$

onde: m_h^0 é a despesa inicial total da h-ésima família; p_{hi}^0 é o preço inicial do item i para família h ; E_{hij}^c refere-se a elasticidade preço da demanda compensada ou Hicksiana; Δp_{hi} é a variação do preço do bem i .

Nesta pesquisa, são usadas duas medidas de VC: sem inclusão do ES, ao assumir $E_{hij}^c = 0$; com ajustamento de demanda, isto é, $E_{hij}^c \neq 0$. Como a VC é calculada para cada família, o seu valor sinaliza o impacto de mudanças na economia no bem-estar dos agentes, que podem repercutir de diferentes formas e intensidades, dependendo da região de localização do domicílio e do estrato de renda. Assim, nesse exercício de microsimulação, admite-se dois cenários de mudanças, que podem ser estimuladas, por exemplo, a partir de alterações nas alíquotas tributárias, em 10% nos preços iniciais do cigarro e das bebidas alcoólicas. Em cada um dos cenários, apenas um dos produtos tem o preço modificado, supondo que os demais se mantenham constantes.

2.4 Base de dados e descrição das variáveis

A estimação de um sistema de demanda requer obrigatoriamente informações sobre preços e despesas dos consumidores para cada categoria de produto. Assim, a Pesquisa de Orçamentos Familiares (POF) 2008-2009 do Instituto Brasileiro de Geografia e Estatística (IBGE) é a principal fonte de dados utilizada para a construção dos resultados. A POF foi realizada entre maio de 2008 e maio de 2009 objetivando traçar a composição da cesta de compras de bens e serviços, estrutura dos rendimentos, estado nutricional e condições de vida das famílias, tendo representatividade para todo o território nacional com cobertura de áreas urbanas e rurais. De acordo com o IBGE (2010), a POF define as estruturas de ponderação utilizadas para o cálculo dos índices de preços ao consumidor, como o Índice de Preço ao Consumidor Amplo (IPCA), servindo também para subsidiar políticas públicas em diversas áreas (tais como social, de saúde e de moradia) e como instrumento estratégico para a definição de investimentos do setor privado.

A POF é composta por sete questionários (POF 1 à POF 7), incluindo além das despesas individuais e coletivas, informações sobre os atributos dos moradores e do domicílio, trabalho e rendimento individual, avaliação das condições de vida e um bloco de perguntas acerca do consumo alimentar pessoal. O período de referência fixado para o deflacionamento das despesas e do rendimento é 15 de janeiro de 2009. A sua amostra inicial é composta por 55.970 unidades domiciliares, com 99,8% apresentando uma unidade

de consumo e 95% com apenas uma única família. Em termos populacionais, o número total de domicílios equivale a 57,69 milhões.

O presente trabalho utiliza nove categorias de produtos, sendo os preços das mercadorias de cada grupo estabelecidos com base na mesma unidade de medida⁹: quilograma (kg) para alimentos e litros (l) para bebidas¹⁰. As categorias examinadas são as seguintes: (1) *Cereais leguminosas e tubérculos* – composto pelas despesas com cereais leguminosas e oleaginosas (arroz, feijão etc.) e tubérculos e raízes (como batata inglesa, cenoura e mandioca); (2) *Massas e panificados* – composto pelos dispêndios com macarrão, farinha de trigo, pão, farinha de mandioca, massas, entre outros; (3) *Frutas e verduras* – despesas com tomate, cebola, alface, banana, laranja, maçã, entre outras frutas e verduras; (4) *Carnes* – nesse grupo estão os gastos com pescados, carne bovina, suína, de aves e de outras espécies; (5) *Leite e derivados* – os dispêndios com leite em geral e produtos como queijo estão inclusos nesse grupo; (6) *Bebidas não-alcoólicas* – composto por refrigerantes, café, sucos, infusões e outras bebidas não-alcoólicas; (7) *Bebidas alcoólicas* – nesse grupo as despesas com cerveja, chope, uísque, champanhe, cachaça, vinho e outras bebidas com teor alcoólico estão incluídas¹¹; (8) *Cigarro* – formado por gastos em nível de domicílio com cigarro; (9) Outros produtos alimentares – formado pelas despesas com produtos alimentares preparados (como lasanha, sanduíche, pizza etc.), enlatados, óleos em geral e açúcares.

Na construção das tabelas e figuras, os nomes dos citados grupos de produtos estão representados pelo primeiro nome dos itens. A escolha dessas categorias é realizada no intuito de ampliar o vetor de consumo das famílias, além da consideração das despesas de interesse com bebidas alcoólicas e cigarro, enquadradas em dispêndios não-duráveis com alimentos, bebidas e artigos de fumo. Como a POF apenas dispõe do valor consumido com cigarro, torna-se necessário capturar o preço desse produto ao longo do território nacional. Para isso, a Pesquisa de Tabagismo da PNAD 2008 é utilizada para computar o preço do cigarro a partir da construção de um coorte com informações para dez quantis de renda domiciliar, 27 unidades federativas e zona de localização da residência (urbana ou rural) para imputação dessa variável nas famílias presentes na POF.

Para evitar problemas de despesas com elevada concentração em um único grupo de produtos, expurgou-se da análise as famílias com esse tipo de observações, por apresentar potenciais problemas informacionais sobre o vetor de consumo. A regra utilizada detecta

⁹ Para as famílias, de uma forma geral, que não computaram valores para os itens, imputou-se preços médios com diferenciação entre as 27 unidades federativas e 10 estratos de renda.

¹⁰ No caso do cigarro, como não houve agregação com outros itens, o preço está medido por unidade do produto.

¹¹ De acordo com os dados da POF 2008-2009, em média, o grupo de bebidas alcoólicas é composto por: cerveja=77,4%; vinho e champanhe=5,6%; aguardente=7,9%; outros destilados=2,1%; demais bebidas alcoólicas=7%. Em termos específicos, no Nordeste o aguardente apresenta um peso bem maior do que a média nacional, em torno de 18%.

os domicílios na cauda do extremo positivo da distribuição dos *shares* de gastos com cada categoria, onde as unidades com *shares* superiores ao valor da média mais duas vezes o desvio-padrão da parcela gasta com o item i foi excluída da análise. Assim, a amostra de 52.007 famílias com dados de despesas alimentares, de bebidas e de cigarro presentes na POF, após o controle para *outliers* e indisponibilidade de informações para alguma variável explicativa, reduziu para 37.671 observações. É possível notar antes do processo de controle, a existência de famílias no banco de dados que informaram 100% de gastos em uma única categoria de produtos. Após a regra de detecção de *outliers*, o vetor de consumo para famílias com despesas positivas tem mais de um grupo de bens. O grupo de carnes, massas e panificados, leite e derivados e bebidas não-alcoólicas possuem, respectivamente, os maiores pesos na cesta de consumo das famílias, no qual esse ordenamento é mantido para os dois tipos de amostras (com e sem observações atípicas).

Além dos controles mencionados, outro problema comumente encontrado nos dados sobre pesquisas de orçamento familiar diz respeito ao número excessivo de despesas nulas em cada grupos de itens (como destacado anteriormente), principalmente para bebidas alcoólicas e cigarro que exibem um elevado percentual de famílias com despesas zeradas (ver Tabela 2.2). Como o registro de gastos na POF é feito ao longo de apenas uma semana para cada domicílio, pode existir situações em que nesse período específico os produtos não foram comprados por questões atreladas às preferências dos consumidores naquele curto intervalo de tempo, por indisponibilidade momentânea de mercadorias no mercado local, por já estarem no estoque da família etc.

Destaca-se que o vetor \mathbf{Z} é composto por três dimensões de variáveis: (1) atributos do morador de referência (chefe) – anos de estudo, idade (em nível e ao quadrado) e gênero; (2) características do domicílio – proporção de moradores analfabetos com mais de 21 anos de idade, número de crianças em duas faixas (de 0 a 11 e de 12 a 17 anos de idade); (3) localização do domicílio – zona urbana e macrorregiões (Norte, Nordeste, Sudeste, Sul e Centro-Oeste). Por sua vez, o vetor \mathbf{J} é formado pelas variáveis pertencentes a \mathbf{Z} , com a inclusão do status de ocupação do chefe, proporção de moradores ocupados no mercado de trabalho e *dummies* sazonais relativas a três períodos festivos (Carnaval, São João e Natal).

De acordo com a Tabela 2.2, que exhibe as estatísticas descritivas das variáveis, a média do logaritmo da despesa total anualizada nos nove itens de produtos é cerca de 8,1 (isto é, R\$ 3.294,47 ao ano ou R\$ 274,54 ao mês), o que representa mensalmente em torno de 60% do salário mínimo em vigor no mês de fevereiro de 2009. Carnes e bebidas alcoólicas apresentam os maiores preços, respectivamente, 1,9 ($e^{\ln p^4} = \text{R\$}6,70$ por kg) e 1,51 ($e^{\ln p^4} = \text{R\$}4,53$ por litro), ao passo que o valor unitário do cigarro apresenta o menor preço, aproximadamente R\$ 0,13 por unidade ($= e^{\ln p^8}$) ou R\$ 2,60 por maço. O morador de referência possui em média 47 anos de idade e 6,9 anos de estudo (ensino fundamental

incompleto), com a grande maioria sendo do sexo masculino (68%) e com alguma ocupação no mercado de trabalho (74%). A maior parte dos domicílios não apresenta frequência de crianças nas duas faixas etárias (de 0 a 11 e de 12 a 17 anos de idade) e, em termos da localização dos domicílios, 85% deles se situam em áreas urbanas, sobretudo nas regiões Sudeste e Nordeste.

Tabela 2.2 – Estatísticas descritivas das variáveis usadas na pesquisa e a proporção de despesas zero por categoria de produto

Sigla	Descrição das variáveis	Média	Desvio-padrão	Gastos zero (%)
w1	Parcela gasta com cereais e tubérculos	0,0837	0,0990	33,7
w2	Parcela gasta com massas e panificados	0,1667	0,1143	5,3
w3	Parcela gasta com frutas e verduras	0,0781	0,0868	29,5
w4	Parcela gasta com carnes	0,2942	0,1901	10,2
w5	Parcela gasta com leite e derivados	0,1122	0,1027	20,1
w6	Parcela gasta com bebidas não-alcoólicas	0,0955	0,0895	20,2
w7	Parcela gasta com bebidas alcoólicas	0,0255	0,0631	80,1
w8	Parcela gasta com cigarro	0,0437	0,1139	77,2
w9	Parcela gasta com outros alimentos	0,1004	0,0984	24,9
lm	Despesa total nos itens avaliados*	8,0958	0,8091	
lnp1	Preço de cereais e tubérculos por <i>kg</i> *	0,7383	0,3235	
lnp2	Preço de massas e panificados por <i>kg</i> *	1,4550	0,4217	
lnp3	Preço de frutas e verduras por <i>kg</i> *	0,7522	0,4198	
lnp4	Preço de carnes por <i>kg</i> *	1,9003	0,3841	
lnp5	Preço de leite e derivados por <i>kg</i> *	1,2790	0,7493	
lnp6	Preço de bebidas não-alcoólicas por <i>l</i> *	1,2509	0,9037	
lnp7	Preço de bebidas alcoólicas por <i>l</i> *	1,5062	0,1747	
lnp8	Preço de cigarro por unidade*	-2,0351	0,1674	
lnp9	Preço de outros alimentos por <i>kg</i> *	1,4494	0,7735	
Morador de referência				
esc _C	Anos de estudo – chefe	6,8994	4,5725	
idade _C	Idade – chefe	47,2243	15,1907	
idade _C ²	Idade ² (dividido por 100) – chefe	24,6089	15,6571	
mulher _C	Mulher – chefe	0,3239	0,4680	
ocupacao _C	Status de ocupação – chefe	0,7392	0,4391	
Atributos do domicílio				
analf _D	Taxa de analfabetos (>21 anos)	0,0698	0,1932	
criancas _{1D}	Crianças entre 0 e 11 anos	0,6841	0,9699	
criancas _{2D}	Crianças entre 12 e 17 anos	0,3930	0,6936	
ocupacao _D	Taxa de ocupados	0,4588	0,3037	
Localização do domicílio				
urbana	Zona Urbana	0,8537	0,3534	
reg1	Região Norte	0,0762	0,2654	
reg2	Região Nordeste	0,2798	0,4489	
reg4	Região Sul	0,1556	0,3624	
reg5	Região Centro-Oeste	0,0642	0,2452	
Dummies sazonais				
carnaval	Mês do Carnaval	0,0781	0,2683	
juninas	Mês do São João	0,1021	0,3028	
natal	Mês do Natal	0,0719	0,2584	
Observações		37.671		
Observações (fator de expansão)		39.937.704		

Fonte: Elaboração própria a partir dos microdados da POF 2008-2009 e da PNAD 2008. Ressalta-se que todas as despesas foram atualizadas e deflacionadas para janeiro de 2009.

* Variáveis expressas em logaritmo neperiano

2.5 Resultados

Esta seção está dividida em três partes. Na primeira são apresentados brevemente características das famílias consumidoras de cigarro e bebidas alcoólicas e as estimativas do modelo QUAIDS que balizam os cálculos das medidas de interesse. A segunda parte exhibe as elasticidades dispêndio e preços (própria e cruzada) para o vetor de produtos. E ao final desta seção, a variação no bem-estar das famílias em resposta às mudanças nos preços dos dois grupos de itens enfatizados no estudo são discutidos.

2.5.1 Característica das famílias demandantes por cigarro e bebidas alcoólicas e estimativas do modelo QUAIDS ajustado

Como a literatura destaca a importância dos fatores sociodemográficos para a demanda das famílias (RAY, 1983; BLUNDELL; PASHARDES; WEBER, 1993; ZHENG; HENNEBERRY, 2010; POI, 2012), a Tabela 2.3 mostra, inicialmente, características gerais das famílias demandantes ($w_i > 0$) e não-demandantes ($w_i = 0$) por bebidas alcoólicas e cigarro no Brasil. Por essa tabela, nota-se que a maior parte dos coeficientes das diferenças de características entre as médias e as proporções dos grupos considerados são significativas estatisticamente a pelos menos 5% , com exceção das variáveis sobre a zona de residência – para as bebidas alcoólicas –, idade e gênero do morador de referência – para o cigarro.

Tabela 2.3 – Características da amostra por grupos de famílias demandantes ($w_i > 0$) e não-demandantes por cigarro e bebidas alcoólicas no Brasil

Variáveis	Bebidas alcoólicas(7)			Cigarro(8)		
	$w_7 = 0$	$w_7 > 0$	Estat. t	$w_8 = 0$	$w_8 > 0$	Estat. t
	(A)	(B)	(B)-(A)	(C)	(D)	(D)-(C)
Média						
esc_C	6,19	7,24	17,19	6,59	5,60	-18,49
$idade_C$	46,72	45,38	-7,02	46,46	46,58	0,70
$criancas_{1D}$	0,77	0,68	-6,73	0,74	0,79	3,15
$criancas_{2D}$	0,43	0,39	-3,72	0,41	0,46	5,64
Renda domiciliar <i>per capita</i>	745,79	1061,14	14,75	831,41	694,78	-7,65
Proporção						
$mulher_C$	34,07%	24,33%	-16,52	32,49%	31,83%	-1,12
$ocupacao_C$	72,77%	80,85%	14,86	73,87%	75,41%	2,85
$analf_D$	8,93%	5,68%	-13,69	7,95%	9,79%	6,82
urbana	78,80%	78,04%	-1,37	79,35%	76,20%	-5,98
Participação em programas sociais	21,57%	15,28%	-12,61	19,62%	23,44%	7,32

Fonte: Elaboração própria a partir dos microdados da POF 2008-2009.

O aspecto central do perfil dos consumidores dos dois produtos reportados na Tabela 2.3 diz respeito às diferentes direções das características socioeconômicas entre os seus demandantes e não-demandantes. Por um lado, as famílias com registro de gastos com bebidas alcoólicas dispõem de um melhor contexto socioeconômico – mais anos de estudo do chefe da família, maior nível de renda domiciliar *per capita*, menor proporção de moradores

com 21 anos ou mais de idade analfabetos e uma menor proporção de participantes em programas sociais – do que aquelas sem registro de consumo de álcool. Por outro, os consumidores de cigarro têm, em média, menos escolaridade, menor rendimento, mais analfabetos e uma maior taxa de beneficiários de programas sociais quando comparados com as unidades não-demandantes do produto. Destarte, a identificação dessas características para os gastos com bebidas e cigarro revelam que a demanda por produtos não saudáveis tende a não responder de forma idêntica a choques nos preços e na renda.

Além dos pontos destacados acima, a questão locacional do domicílio pode implicar em diferentes comportamentos no perfil dos consumidores. Conforme a POF 2008-2009, o desenho do consumo entre as famílias situadas em diferentes regiões possui algumas similaridades entre eixos regionais, pois os valores despendidos nos eixos Centro-Sul são mais similares do que os registrados no Norte-Nordeste. Tais aspectos sugerem, por exemplo, que políticas tributárias específicas para determinados produtos podem afetar o bem-estar dos agentes de uma região com mais intensidade do que em outras.

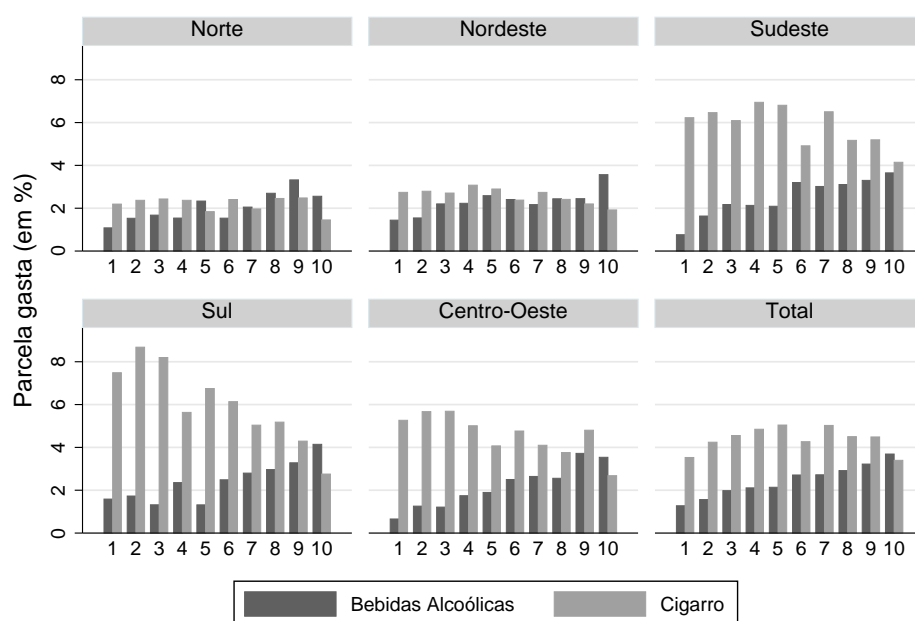
Dentre os nove grupos de produtos, em média, a parcela gasta com bebidas alcoólicas e cigarro registra, independentemente da região de localização, a menor participação no orçamento domiciliar restrito ao vetor composto por essas nove categorias – quando se considera a amostra total. Quanto ao baixo valor relativo da parcela gasta com esses produtos, salienta-se que eles não são consumidos necessariamente por todos os membros da família, diferentemente do consumo de outros produtos – como carnes ou leite –, bem como os dados da POF para esses grupos de itens registram um excessivo número de despesas nulas.

Tendo em apreço as 6.680 observações com valores não-nulos de despesas com bebidas alcoólicas no período e na amostra final, este grupo de produtos mostra-se o segundo item no vetor de consumo com maior peso no orçamento, computando um dispêndio, em média, de R\$ 868,84 ao ano. Restringindo também para as 8.190 famílias com registro de algum consumo de cigarro, este produto, a exemplo do ocorrido com as bebidas alcoólicas, passa para o segundo posto no ordenamento das maiores despesas da família dentre as nove categorias de produtos – com gastos anuais superiores a R\$ 695,00. Ademais, para as 2.126 famílias que possuem consumo em ambas categorias, o cigarro e bebidas com teor alcoólico ocupam, em média, o segundo e o terceiro posto no ordenamento do dispêndio destas famílias com, respectivamente, 14,1% e 13,1% da despesa total.

A Figura 2.2 exibe a parcela gasta, em média, com bebidas alcoólicas e cigarro por decil da renda domiciliar *per capita* para as macrorregiões do país e no total. Em termos globais, fica claro que o peso dos gastos para os diferentes decis de renda com esses dois grupos de produtos, em especial para o cigarro, nas regiões Norte e Nordeste é inferior ao observado pelas famílias residentes nas demais regiões. O comportamento

dos gastos com bebidas alcoólicas no cômputo total aponta que quanto maior o nível de rendimento familiar *per capita* maior o peso das bebidas no orçamento em todos os níveis regionais. Quanto ao consumo do cigarro, as famílias de menor nível de renda nos eixos Centro-Sul dispõem de uma parcela gasta bem acima da média nacional. A título de exemplo, enquanto que famílias do primeiro decil de renda no Sul do país registram um valor da parcela gasta com cigarro duas vezes maior do que a média nacional para esse grupo de renda, o *share* desse item é inferior a média nacional equivalente para as famílias do último decil.

Figura 2.2 – Parcela gasta, em média, com bebidas alcoólicas e cigarro por decil da renda domiciliar *per capita* nas macrorregiões brasileiras



Fonte: Elaboração própria a partir dos microdados da POF 2008-2009.

Dado o escopo da pesquisa em analisar a demanda das bebidas alcoólicas e cigarro, sem desconsiderar outros produtos alimentares da cesta de consumo, a Tabela 2.4 apresenta os parâmetros que calibram o cálculo das elasticidades e da variação compensatória por meio da abordagem QUAIDS com as correções para consumo censurado e endogeneidade das despesas totais¹². Destaca-se que as interpretações da abordagem QUAIDS, de acordo com Banks, Blundell & Lewbel (1997), devem ser realizadas em cima das elasticidades, visto as complexas relações entre as diferentes covariadas no sistema de equações não-lineares de demanda.

¹² Cabe ressaltar que as estimativas do modelo QUAIDS sem o citado ajustamento estão disponíveis na Tabela A.3 no Apêndice. Baseado nelas, as elasticidades dispêndio e preços são desenvolvidas para fins comparativos com os valores auferidos pelo o modelo ajustado.

Tabela 2.4 – Parâmetros do QAIDS ajustado aos valores censurados e endogeneidade das despesas totais

Parâmetros	Equações das composições dos grupos de produtos							
	Cereais(1)	Massas(2)	Frutas(3)	Carnes(4)	Leite(5)	S/Álcool(6)	Álcool(7)	Cigarro(8)
α_i	0,116345***	0,356133***	0,094695***	0,314681***	0,130451***	0,190730***	-0,017268	-0,207281***
β_i	0,017714*	-0,139649***	-0,002438	-0,071634***	-0,041341***	-0,048090***	0,189955***	-0,150298***
γ_{i1}	0,038572***	-0,008298***	0,001876	-0,017068***	-0,001328	0,009883***	0,017684***	-0,012803***
γ_{i2}	-0,008298***	-0,012321***	0,003573	-0,038322***	-0,006760***	-0,011484***	0,045795***	-0,009943*
γ_{i3}	0,001876	0,003573	-0,007422**	0,000032	0,001322	0,004878***	0,00613	-0,016956***
γ_{i4}	-0,017068***	-0,038322***	0,000032	0,003674***	-0,006946***	-0,008979***	0,031043***	-0,010089**
γ_{i5}	-0,001328	-0,006760***	0,001322	-0,006946***	-0,019086***	-0,000001	0,022929***	-0,015489***
γ_{i6}	0,009883***	-0,011484***	0,004878**	-0,008979***	-0,000001	-0,023148***	0,013733***	-0,004665
γ_{i7}	0,017684***	0,045795**	0,00613	0,031043***	0,022929***	0,013733***	-0,185942***	0,085720***
γ_{i8}	-0,012803***	-0,009943*	-0,016956***	-0,010089**	-0,015489***	-0,004665	0,085720***	-0,091467***
λ_i	0,000235	0,025569***	-0,004055***	0,006039***	0,013210***	0,012938***	-0,032636***	0,010761**
$\eta_i(\text{idade}_C)$	-0,000701***	0,000429***	0,000259***	-0,001275***	0,000977***	0,000204**	-0,000960***	-0,001053***
$\eta_i(\text{idade}_C^2)$	0,000377***	-0,000455***	-0,000133	0,000858***	-0,000705***	-0,000287**	0,000299	0,002466***
$\eta_i(\text{idade}_C^2)$	0,000247**	0,000507***	0,000291**	-0,000811***	0,000758***	0,000205*	-0,000576*	-0,002502***
$\eta_i(\text{mulher}_C)$	-0,000167	0,000439	0,000304	-0,004293***	0,001378*	-0,000007	-0,001746	0,005372**
$\eta_i(\text{analf}_D)$	0,008199***	-0,002815	-0,004508**	0,008951**	-0,001911	0,001441	-0,006373	-0,007041
$\eta_i(\text{criancas}_{1D})$	0,000005	0,003277***	-0,000923***	-0,000155	0,001664***	-0,001342***	-0,002616***	-0,000518
$\eta_i(\text{criancas}_{2D})$	0,000793*	0,004474***	-0,002038***	0,001115	-0,002560***	0,000484	-0,001972*	-0,002770**
$\eta_i(\text{urbana})$	-0,004200***	0,003265***	-0,000028	-0,005190***	0,002644***	-0,000003	-0,001992	0,007854***
$\eta_i(\text{reg1})$	-0,010542***	0,011665***	-0,002429**	0,044607***	-0,009270***	0,001099	-0,010903***	-0,025818***
$\eta_i(\text{reg2})$	-0,004111***	0,017089***	0,000654	0,025407***	-0,000312	-0,006639***	-0,008066***	-0,027479***
$\eta_i(\text{reg4})$	-0,005478***	0,005238***	-0,003050***	0,006411***	-0,003182***	0,001163	-0,002425	0,001878
$\eta_i(\text{reg5})$	0,008608***	-0,006607***	0,003096***	0,010252***	-0,005947***	0,001572	0,000604	-0,015812***
$\rho(\text{esc}_C)$	-0,00046	-0,00046	-0,00046	-0,00046	-0,00046	-0,00046	-0,00046	-0,00046
$\rho(\text{idade}_C)$	-0,002204	-0,002204	-0,002204	-0,002204	-0,002204	-0,002204	-0,002204	-0,002204
$\rho(\text{idade}_C^2)$	0,002952	0,002952	0,002952	0,002952	0,002952	0,002952	0,002952	0,002952
$\rho(\text{mulher}_C)$	-0,068158***	-0,068158***	-0,068158***	-0,068158***	-0,068158***	-0,068158***	-0,068158***	-0,068158***
$\rho(\text{analf}_D)$	0,085706**	0,085706**	0,085706**	0,085706**	0,085706**	0,085706**	0,085706**	0,085706**
$\rho(\text{criancas}_{1D})$	0,009715	0,009715	0,009715	0,009715	0,009715	0,009715	0,009715	0,009715
$\rho(\text{criancas}_{2D})$	0,038694***	0,038694***	0,038694***	0,038694***	0,038694***	0,038694***	0,038694***	0,038694***
$\rho(\text{urbana})$	-0,110541***	-0,110541***	-0,110541***	-0,110541***	-0,110541***	-0,110541***	-0,110541***	-0,110541***
$\rho(\text{reg1})$	-0,267942***	-0,267942***	-0,267942***	-0,267942***	-0,267942***	-0,267942***	-0,267942***	-0,267942***
$\rho(\text{reg2})$	-0,226213***	-0,226213***	-0,226213***	-0,226213***	-0,226213***	-0,226213***	-0,226213***	-0,226213***
$\rho(\text{reg4})$	-0,160269***	-0,160269***	-0,160269***	-0,160269***	-0,160269***	-0,160269***	-0,160269***	-0,160269***
$\rho(\text{reg5})$	0,097244***	0,097244***	0,097244***	0,097244***	0,097244***	0,097244***	0,097244***	0,097244***
τ_i	0,018064*	0,051561**	-0,048871***	0,301904***	0,035493***	0,01081	0,118788***	0,212380***
ϑ_i	-0,003940**	-0,001082	-0,020976***	0,042331***	-0,020539***	-0,020639***	-0,004662**	-0,004238***
α_0	4,43	4,43	4,43	4,43	4,43	4,43	4,43	4,43
R^2	0,52	0,71	0,50	0,73	0,58	0,59	0,21	0,17
N	37.671	37.671	37.671	37.671	37.671	37.671	37.671	37.671
P	39.939.352	39.939.352	39.939.352	39.939.352	39.939.352	39.939.352	39.939.352	39.939.352

Fonte: Elaboração própria a partir dos microdados da POF 2008-2009 e da PNAD 2008.

Nota: A equação omitida do sistema refere-se ao grupo de outros alimentos, onde os parâmetros são obtidos através das condições de aditividade, simetria de Slutsky e homogeneidade. *p-valor<0,1; **p-valor<0,05; ***p-valor<0,01.

Legenda: S/Álcool = bebidas sem teor alcoólico.

De uma forma geral, os parâmetros na Tabela 2.4, além de possuírem significância estatística, mostram que o consumo de uma dada categoria de produtos está relacionado com os diferentes vetores de variáveis associadas aos preços, renda e fatores conjunturais (socioeconômicos e locacionais). O grau de ajuste do modelo, capturado pelo R^2 , demonstra que as covariadas apresentam um grau de explicação que varia de 0,17 a 0,73, sendo o modelo para a demanda por cigarro o que tem menor ajustamento e o de carnes com o maior poder preditivo. O parâmetro λ associado ao termo quadrático da despesa total ratifica para todas as equações, com exceção para o consumo de cereais, a existência de uma não-linearidade entre o gasto total e o consumo de um dado item.

As Tabelas A.1 e A.2 no Apêndice reportam as estimativas que subsidiaram os cálculos da FDA e da FDP, que ajustam os parâmetros do modelo QUAIDS¹³. No que tange às estimativas atreladas a FDP na Tabela 2.4 – relacionada com o processo de correção dos valores censurados – e ao resíduo da equação de segundo estágio¹⁴ – usada para ajuste da endogeneidade da despesa total –, elas são estatisticamente significativas para quase totalidade das equações. Esses coeficientes revelam a valia do controle de possíveis vieses motivados pelos problemas empíricos e intrínsecos ao conjunto de dados.

Os resultados expostos pelos parâmetros do sistema de demanda indicam a importância da renda e dos preços (do próprio grupo de produtos e do demais) para a determinação da parcela gasta com cada categoria de itens alimentares, bebidas e cigarro. As relações dos fatores sociodemográficos e locacionais também mostram-se relevantes para a calibração do sistema de demanda, devido ao número expressivo de coeficientes estatisticamente diferentes de zero. Cabe ressaltar mais uma vez, que os parâmetros presentes na Tabela 2.4 não possuem interpretações diretas, por isso que as avaliações direcionadas para o escopo do trabalho são explicitadas com mais detalhes nas próximas duas subseções.

2.5.2 Elasticidades

Para minimizar os efeitos de valores atípicos, as análises são feitas para a mediana das elasticidades dispêndio e preços da demanda. Apenas para fins comparativos, os grupos de produtos que não fazem parte dos objetivos centrais desta pesquisa são analisados de forma mais ampla. As elasticidades-dispêndio por decil de renda domiciliar *per capita* e por níveis macrorregionais são reportadas na Tabela 2.5, admitindo um intervalo de confiança de 95% para o valor da mediana. Os limites mínimos e máximos do intervalo

¹³ A matriz de correlação dos termos de erros na Tabela A.2 no Apêndice evidencia, de uma forma geral, que os erros estocásticos de uma dada expressão do indicador de consumo possui relações significativamente estatísticas com os erros de outras equações, ratificando a importância da utilização do modelo probit multivariado para o caso em questão.

¹⁴ A Figura A.2 no Apêndice apresenta os valores médios dos resíduos do estágio 2 por decil de renda, onde nota-se que quanto menor o nível de renda da família maior são os resíduos da regressão.

estão, respectivamente, entre parênteses e colchetes.

Tabela 2.5 – Elasticidade-dispêndio da demanda por decil de renda domiciliar *per capita* e por níveis regionais no Brasil 2008-2009

Por	Grupo de itens								
	Cereais (1)	Massas (2)	Frutas (3)	Carnes (4)	Leite (5)	S/Álcool (6)	Álcool (7)	Cigarro (8)	Outros (9)
<i>Decil</i>									
1°	0,99959 (0,99956) [0,99965]	0,80751 (0,80748) [0,80792]	0,81600 (0,81541) [0,81655]	0,96269 (0,96257) [0,96278]	0,96415 (0,96399) [0,96421]	0,85209 (0,85161) [0,85239]	1,06198 (1,06198) [1,06215]	0,76503 (0,76315) [0,76751]	1,46049 (1,46049) [1,46049]
2°	0,99446 (0,99443) [0,99449]	0,81346 (0,81313) [0,81429]	0,82905 (0,82903) [0,82905]	0,94952 (0,94952) [0,94959]	0,97816 (0,97795) [0,97823]	0,89718 (0,89654) [0,8973]	1,05622 (1,05622) [1,05758]	0,83207 (0,83148) [0,83207]	1,36185 (1,36185) [1,37457]
3°	0,99058 (0,99058) [0,99063]	0,78601 (0,78519) [0,78609]	0,84277 (0,84229) [0,8428]	0,93469 (0,93456) [0,93508]	0,97848 (0,97819) [0,9787]	0,90781 (0,90777) [0,90788]	1,06542 (1,06542) [1,06661]	0,86651 (0,86538) [0,86666]	1,72551 (1,59742) [1,72551]
4°	0,98932 (0,98922) [0,98935]	0,76996 (0,76988) [0,77026]	0,84932 (0,84827) [0,84972]	0,92885 (0,92884) [0,92885]	0,98351 (0,98343) [0,98357]	0,91803 (0,91776) [0,91803]	1,0755 (1,07528) [1,0755]	0,87736 (0,87667) [0,87754]	1,25064 (1,25064) [1,25064]
5°	0,98731 (0,98731) [0,98731]	0,75877 (0,75877) [0,75895]	0,8595 (0,85915) [0,85963]	0,91698 (0,91694) [0,91698]	0,9897 (0,98956) [0,9897]	0,92091 (0,92089) [0,92101]	1,07557 (1,07541) [1,07569]	0,88811 (0,8873) [0,88822]	1,64146 (1,64146) [1,65343]
6°	0,98796 (0,9879) [0,98798]	0,74743 (0,74666) [0,74757]	0,8594 (0,85931) [0,85972]	0,91635 (0,91627) [0,91656]	0,9905 (0,98962) [0,99099]	0,93094 (0,93094) [0,93119]	1,07623 (1,07623) [1,07644]	0,88851 (0,88735) [0,88851]	1,26383 (1,26383) [1,26383]
7°	0,97598 (0,97596) [0,97612]	0,7288 (0,7288) [0,7288]	0,86037 (0,86001) [0,8604]	0,90599 (0,90592) [0,90615]	0,99402 (0,9938) [0,99406]	0,92895 (0,92895) [0,92914]	1,08291 (1,08291) [1,08291]	0,92023 (0,91906) [0,92023]	1,34489 (1,34489) [1,39835]
8°	0,9714 (0,97116) [0,97144]	0,73917 (0,73895) [0,73917]	0,86869 (0,86869) [0,86869]	0,89583 (0,8957) [0,89607]	1,01122 (1,0111) [1,01175]	0,94132 (0,94121) [0,94142]	1,08079 (1,08079) [1,08079]	0,91597 (0,91519) [0,91633]	1,48327 (1,4763) [1,48327]
9°	0,95484 (0,95466) [0,95512]	0,73865 (0,73865) [0,73877]	0,87599 (0,87599) [0,87611]	0,87892 (0,87892) [0,87892]	1,01967 (1,01939) [1,01967]	0,95167 (0,95145) [0,95167]	1,07588 (1,07528) [1,07606]	0,92631 (0,92631) [0,92631]	1,33587 (1,33587) [1,33757]
10°	0,91343 (0,91337) [0,9142]	0,76361 (0,76349) [0,76362]	0,88419 (0,88347) [0,88433]	0,84769 (0,84761) [0,84773]	1,04861 (1,04834) [1,04903]	0,96004 (0,96000) [0,96009]	1,09028 (1,08865) [1,09035]	0,92918 (0,92854) [0,92987]	1,35153 (1,35153) [1,35153]
<i>Região</i>									
NO	0,92838 (0,92834) [0,92846]	0,8421 (0,8419) [0,84212]	0,76707 (0,76633) [0,76715]	1,00863 (1,00863) [1,00864]	0,96434 (0,96411) [0,96456]	0,95997 (0,95994) [0,96012]	1,03303 (1,03303) [1,03312]	0,7517 (0,74938) [0,75546]	1,20439 (1,20439) [1,20439]
NE	0,9841 (0,98405) [0,98411]	0,82595 (0,82586) [0,82597]	0,86156 (0,86138) [0,86185]	1,00616 (0,96161) [0,96165]	1,00375 (1,00356) [1,00376]	0,8624 (0,86234) [0,86243]	1,06272 (1,06263) [1,06272]	0,7986 (0,79782) [0,79924]	1,37916 (1,37457) [1,40276]
SE	0,98664 (0,98663) [0,98666]	0,70445 (0,70417) [0,70476]	0,88061 (0,88059) [0,88073]	0,86096 (0,86094) [0,86104]	1,00781 (1,00747) [1,00791]	0,9357 (0,93554) [0,93586]	1,09696 (1,09692) [1,09726]	0,92433 (0,92429) [0,92434]	1,4658 (1,4658) [1,47302]
SU	0,94136 (0,94135) [0,94148]	0,80393 (0,80393) [0,80398]	0,7993 (0,79883) [0,79976]	0,88868 (0,88857) [0,88868]	1,0218 (1,02162) [1,02194]	0,97832 (0,97800) [0,97836]	1,0619 (1,06185) [1,0619]	0,91058 (0,91057) [0,91117]	1,28673 (1,28281) [1,28673]
CO	1,03586 (1,03584) [1,03588]	0,59218 (0,59136) [0,59366]	0,91789 (0,91766) [0,91809]	0,89797 (0,89774) [0,89801]	0,94779 (0,94769) [0,94812]	0,92727 (0,92717) [0,92729]	1,09369 (1,09369) [1,09396]	0,89383 (0,89284) [0,89397]	1,42618 (1,41598) [1,42618]
Total	0,9804 (0,98035) [0,98041]	0,76374 (0,76368) [0,76382]	0,86124 (0,8611) [0,86126]	0,91049 (0,91047) [0,91049]	1,00182 (1,00173) [1,00186]	0,92963 (0,92955) [0,92966]	1,0773 (1,07709) [1,0773]	0,89554 (0,89546) [0,89557]	1,37457 (1,37289) [1,37916]

Fonte: Elaboração própria a partir dos microdados da POF 2008-2009 e da PNAD 2008. Valores calculados para mediana, com intervalo de confiança de 95% para a mediana ajustados para o fator de expansão amostral. Os valores entre parênteses e entre colchetes referem-se, respectivamente, aos limites mínimos e máximos do intervalo para a mediana. NO = Norte; NE = Nordeste; SE = Sudeste; SU = Sul; CO = Centro-Oeste. S/Álcool = bebidas sem teor alcoólico.

Segundo a Tabela 2.5, todos os produtos são bens normais ($E > 0$), porque aumentos na renda da família implicam em um maior consumo deles. Dentre os produtos

classificados como necessários ($0 < E < 1$), o cigarro é o menos sensível a alteração na despesa total para famílias do menor decil de renda, bem como para as localizadas no Norte e Nordeste do país. Sem contabilizar a categoria residual (outros alimentos), o grupo de produtos com maior elasticidade-dispêndio é o de bebidas alcoólicas para qualquer decil de renda e agrupamento regional, podendo ser tratado como bem supérfluo (ou de luxo), já que variações nas despesas familiares expandem mais do que proporcionalmente o seu consumo. Com ganhos positivos nos seus rendimentos, as famílias, mormente as residentes no Sudeste e Centro-Oeste, ampliam a demanda por bebidas alcoólicas em uma razão maior do que para produtos como cigarro, bebidas não-alcoólicas e frutas e verduras.

Considerando o público-alvo dos programas sociais de transferência de renda, como o Bolsa Família, constata-se que a elasticidade-dispêndio maior que um para o consumo de bebidas alcoólicas é mantida para os menores decis de renda na Tabela 2.5. Adicionalmente, ao controlar a elasticidade-dispêndio para bebidas alcoólicas de unidades com baixa renda e beneficiárias de programas sociais federais chefiadas por homens e mulheres, a demanda é elástica para ambos os tipos de moradores de referência, porém maior para domicílios chefiados por homens (1,13 contra 1,04). Portanto, choques positivos na renda dos mais pobres repercutem relativamente com mais intensidade na demanda de produtos com teor alcoólico do que em itens alimentares, pois um incremento de 10% nos dispêndios das famílias mais pobres resultam em uma elevação de 10,6% no consumo de bebidas alcoólicas.

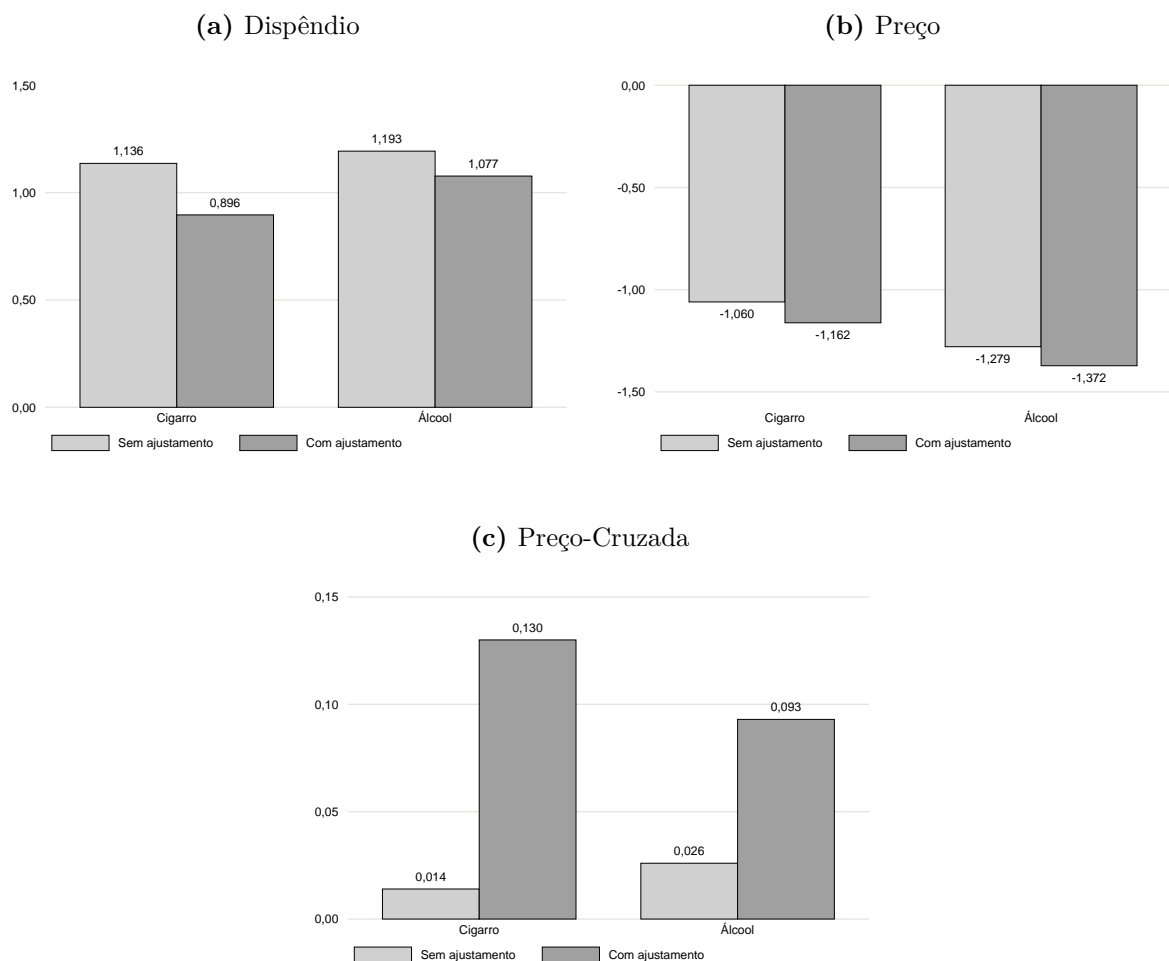
A resposta da demanda à modificações nos dispêndios apresenta uma tendência crescente com o decil de renda domiciliar, exceto para as categorias de massas, cereais e carnes. Para o cigarro, a elasticidade-dispêndio reproduz essa especificidade. No total, um acréscimo de 10% nos gastos familiares aumentam o consumo do produto em cerca de 9%, enquanto que para as famílias do primeiro e do último decil de renda esta alta é de, respectivamente, 7,65% e 9,3%. Logo, a elasticidade para a demanda por cigarro do último decil de renda domiciliar *per capita* é 21% maior do que a do primeiro decil, sendo essa a maior diferença registrada dentre os demais produtos entre os extremos da distribuição dos rendimentos. Além do mais, para os mais pobres, o cigarro possui a menor elasticidade entre todos os itens.

No comparativo com resultados de outros estudos desenvolvidos para o Brasil, conforme apresentado na Tabela 2.1, as estimativas calculadas seguem a tendência de $E_7 > 1$ para bebidas alcoólicas e $0 < E_8 < 1$ para o cigarro. Em Pintos-Payeras (2009) e Yamamoto (2011), os valores para a elasticidade-dispêndio da demanda para bebidas alcoólicas varia de 1,11 a 1,38, por outro lado esta medida para o cigarro situa-se entre 0,23 e 0,63 em Carvalho & Lobao (1998), Menezes, Silveira & Azzoni (2008) e Pintos-Payeras (2009). Cabe realçar que nessas pesquisas as estimativas para as elasticidades são obtidas com diferentes períodos, grupos amostrais e agregação dos dados, como em Carvalho &

Lobao (1998), além do mais os estudos com a abordagem AIDS não corrigem os problemas empíricos enfatizados pela literatura.

A Figura 2.3 reporta as estimativas das elasticidades para os modelos sem e com ajustamento para consumo censurado e endogeneidade das despesas totais, demonstrando que as elasticidades-dispêndio da demanda do cigarro e de bebidas alcoólicas podem ser sobrestimadas pela abordagem QUAIDS sem às devidas correções no modelo empírico. De acordo com a Figura 2.3a, esse viés é maior para o cigarro, já que a taxa de variação das estimativas entre os modelos com e sem ajuste é de cerca de -21%, ao tempo que para as bebidas alcoólicas o valor é de -10% .

Figura 2.3 – Elasticidades estimadas para os modelos sem e com ajustamento para consumo censurado e endogeneidade das despesas totais, bebidas e cigarro no Brasil 2008-2009



Fonte: Elaboração própria a partir dos microdados da POF 2008-2009 e PNAD 2008. Valores calculados para mediana.
 Nota: As elasticidades preço e preço-cruzada correspondem a demanda não-compensada. Os parâmetros que calibram as elasticidades do modelo sem ajustamento estão apresentados na Tabela A.3 no Apêndice. As diferenças entre as elasticidades com e sem ajustamento são estatisticamente diferentes de zero a pelo menos 5% de significância.

Por meio da Figura 2.3, as elasticidades dispêndio e preço têm os mesmos sinais nas duas modelagens para os dois tipos de produtos, mas com magnitudes estatisticamente díspares em todos os casos. Diferentemente da elasticidade-dispêndio, as elasticidades-preço do próprio produto e cruzada estão, em termos absolutos, subestimadas no modelo sem ajustamento. No caso da elasticidade-preço cruzada da demanda as diferenças entre as estimativas são as mais marcantes, indicando a relação de substitutibilidade entre os produtos.

Nessa direção, a Tabela 2.6 exibe a matriz de elasticidades-preço das demandas Marshallianas e Hicksianas para os nove grupos de itens a partir dos parâmetros auferidos pelo modelo com ajustamento. As elasticidades da diagonal principal da matriz explicitam as condições de demanda do produto em relação ao seu próprio preço e os demais coeficientes fora desta diagonal correspondem as relações de demanda de um produto condicionado ao preço de outro.

Tabela 2.6 – Matriz de elasticidades-preço das demandas Marshallianas e Hicksianas no Brasil 2008-2009

Preço→ Grupo↓	Cereais (1)	Massas (2)	Frutas (3)	Carnes (4)	Leite (5)	S/Álcool (6)	Álcool (7)	Cigarro (8)	Outros (9)
Demanda não-compensada ou Marshalliana									
1.Cereais	-0,8196	-0,1174	-0,0155	-0,1686	-0,0003	0,1034	0,1423	-0,0989	-0,2825
2.Massas	-0,0213	-0,8730	0,0276	-0,0942	0,0289	0,0087	0,0615	0,0810	-0,0117
3.Frutas	-0,4653	0,3094	-1,2532	0,0959	-0,2279	-0,0180	0,0450	-0,1577	0,1814
4.Carnes	-0,0448	-0,0877	0,0108	-0,8601	-0,0037	-0,0062	0,0447	-0,0075	-0,0005
5.Leite	-0,0096	0,0155	-0,3439	0,0597	-1,1134	0,0657	0,0667	-0,0326	-0,0072
6.S/Álcool	0,0763	0,0059	0,0581	-0,6526	0,1673	-1,1365	0,0035	0,0803	-0,0135
7.Álcool	0,0442	0,0111	0,0018	0,0181	0,0206	0,0292	-1,3723	0,0931	0,0294
8.Cigarro	-0,0728	0,0146	-0,0238	0,0141	0,5083	-0,0298	0,1296	-1,1620	0,0816
9.Outros	0,1837	-0,1923	0,1962	0,0391	-1,0493	-0,2060	0,0139	0,0581	-0,7772
Demanda compensada ou Hicksiana									
1.Cereais	-0,6719	0,0098	0,0599	0,0583	0,0765	0,1972	0,1587	-0,0863	-0,1971
2.Massas	0,0253	-0,7178	0,0888	0,0665	0,1104	0,0653	0,0701	0,0978	0,0399
3.Frutas	-0,3986	0,4461	-1,1638	0,3369	-0,1421	0,0603	0,0562	-0,1427	0,2806
4.Carnes	-0,0119	0,0502	0,0604	-0,5380	0,0779	0,0501	0,0522	0,0009	0,0651
5.Leite	0,0293	0,1851	-0,2630	0,3264	-0,9910	0,1450	0,0823	-0,0257	0,1056
6.S/Álcool	0,1590	0,1516	0,1292	-0,4153	0,2504	-1,0433	0,0148	0,1019	0,0711
7.Álcool	0,1194	0,1115	0,0575	0,2839	0,1279	0,1400	-1,2320	0,1214	0,1532
8.Cigarro	-0,0216	0,1480	-0,0113	0,2507	0,5916	0,0167	0,1759	-1,0419	0,1598
9.Outros	0,3154	-0,0545	0,3082	0,3740	-0,9427	-0,0742	0,1135	0,1709	-0,5888

Fonte: Elaboração própria a partir dos microdados da POF 2008-2009 e da PNAD 2008. Valores calculados para mediana.
Legenda: S/Álcool = bebidas sem teor alcoólico.

Através da matriz de elasticidade-preço da demanda, disposta na Tabela 2.6, todas as elasticidades-preço dos próprios produtos dispostas na diagonal principal são negativas tanto na demanda Marshalliana quanto na Hicksiana, condizentes com os preceitos teóricos que discorre sobre a relação inversa entre preço e quantidade demandada. Pela demanda não-compensada, a elasticidade-preço para o cigarro (E_{88}^u) é de -1,16 e para álcool (E_{77}^u) de -1,37, em que uma variação de 1% no próprio preço de tais produtos repercute em uma

redução na quantidade consumida em uma razão maior do que uma unidade. Achados similares aos de [Menezes, Silveira & Azzoni \(2008\)](#) e [Pintos-Payeras \(2009\)](#) para o cigarro, que observam uma elasticidade-preço de -2,84 e -1,03, e de -1,11 para a bebida alcoólica em [Pintos-Payeras \(2009\)](#). Desta maneira, a demanda por tais categorias de itens é mais elástica ao próprio preço do que a maior parte dos produtos alimentares e bebidas não-alcoólicas que compõem o vetor de consumo.

Baseado nas estimativas, uma variação positiva de 10% no preço do cigarro, por exemplo, reduz em 11,6% o consumo do produto no âmbito do orçamento domiciliar¹⁵. Tais dados ilustram que políticas tarifárias restritivas podem ter um impacto expressivo no desincentivo da demanda pelos produtos associados com os fatores de risco comportamentais a saúde (tanto para o cigarro quanto para as bebidas alcoólicas). Inclusive, de acordo com [Chaloupka & Grossman \(1996\)](#), tarifas adicionais sobre o cigarro levam a drásticas reduções no seu consumo para os indivíduos mais jovens e na probabilidade do mesmo decidir ser fumante, apontando que políticas tarifárias podem ter certa efetividade no controle deste vício.

Sobre a elasticidade para o cigarro maior que uma unidade calculada para o caso brasileiro, algumas hipóteses podem ser realçadas, a saber: quantitativo relativamente grande de pessoas com registro esporádico de consumo, girando em torno de 20% do número total de indivíduos com a experiência do uso do cigarro, segundo a PNAD 2008 – de modo que para este perfil de usuários as mudanças nos preços podem gerar uma maior sensibilidade na demanda do que para os consumidores com maior frequência de utilização; legislação não-tarifária, como a proibição do uso do cigarro em locais fechados e advertências sobre as consequências do tabagismo a saúde impressas nos maços do produto, pode contribuir com uma mudança nas preferências dos usuários, tornando-os mais susceptíveis a choque nos preços do cigarro – conforme a pesquisa especial de tabagismo da PNAD 2008, por exemplo, mais de 90% dos inqueridos se sentem tocados pelas advertências sobre os riscos do cigarro, pensando inclusive em parar de fumar; demanda por bebida alcoólica incluída no sistema, pois segundo [Gallet & List \(2003\)](#) a elasticidade-preço do cigarro fica maior quando estimada em conjunto com a demanda por álcool.

Nas Figuras 2.4a e 2.4b são mostradas as elasticidades-preço da demanda não-compensada por decil de renda domiciliar *per capita* para as bebidas alcoólicas e cigarro. Por meio dessa ilustração, verifica-se que a demanda elástica dos dois produtos é identificada para diferentes níveis econômicos das famílias. Todavia, o consumo de bebidas com teor alcoólico é mais sensível ao preço para as famílias mais ricas do que para as mais pobres,

¹⁵ Ressalta-se que quanto maior o peso das despesas com o cigarro no orçamento familiar, mais inelástica a demanda por este produto à mudança no preço. Esse fato parece condizente com as definições de [Becker & Murphy \(1988\)](#) sobre bens de vício, uma vez que o consumo destes produtos se relacionam com o seu estoque de consumo pretérito.

possivelmente reverberando os distintos tipos de bebidas consumidos entre elas. Já que famílias com menor rendimento registram um maior parcela de despesas com bebidas alcoólicas com baixo valor de mercado (como aguardente) quando se compara com as bebidas consumidas pelos domicílios com maior nível de renda, que adquirem itens com maior preço (como vinho, uísque etc.). Em relação a Figura 2.4b, os decis mais baixos de renda têm uma maior sensibilidade no consumo do cigarro a mudanças no seu preço do que para as famílias com maior nível econômico, mas essas diferenças são menores do que para as bebidas alcoólicas. Esta última constatação, pode ser fruto relativamente da menor amplitude de preços e variedade de cigarro.

Pela tendência central das estimativas, bebidas alcoólicas e cigarro são bens substitutos ($E_{78}^u = 0,09$ e $E_{87}^u = 0,13$). A quantidade demandada de cigarro, entretanto, tem uma resposta ligeiramente maior à modificações nos preços de bebidas alcoólicas do que o inverso. Em Pintos-Payeras (2009), esses produtos mantêm uma baixa relação de complementariedade ($E_{78}^u = -0,04$ e $E_{87}^u = -0,016$). Logo, tendo em vista que a incorporação de ajustes no sistema de demanda para questões como valores censurados de consumo (algo que merece atenção, principalmente para as informações de gastos com cigarro e bebidas alcoólicas) e maior representatividade amostral, os achados deste trabalho acompanham indicações de outros estudos na literatura empírica internacional¹⁶ como Fan, Cramer & Wailes (1994), Goel & Morey (1995) e Decker & Schwartz (2000).

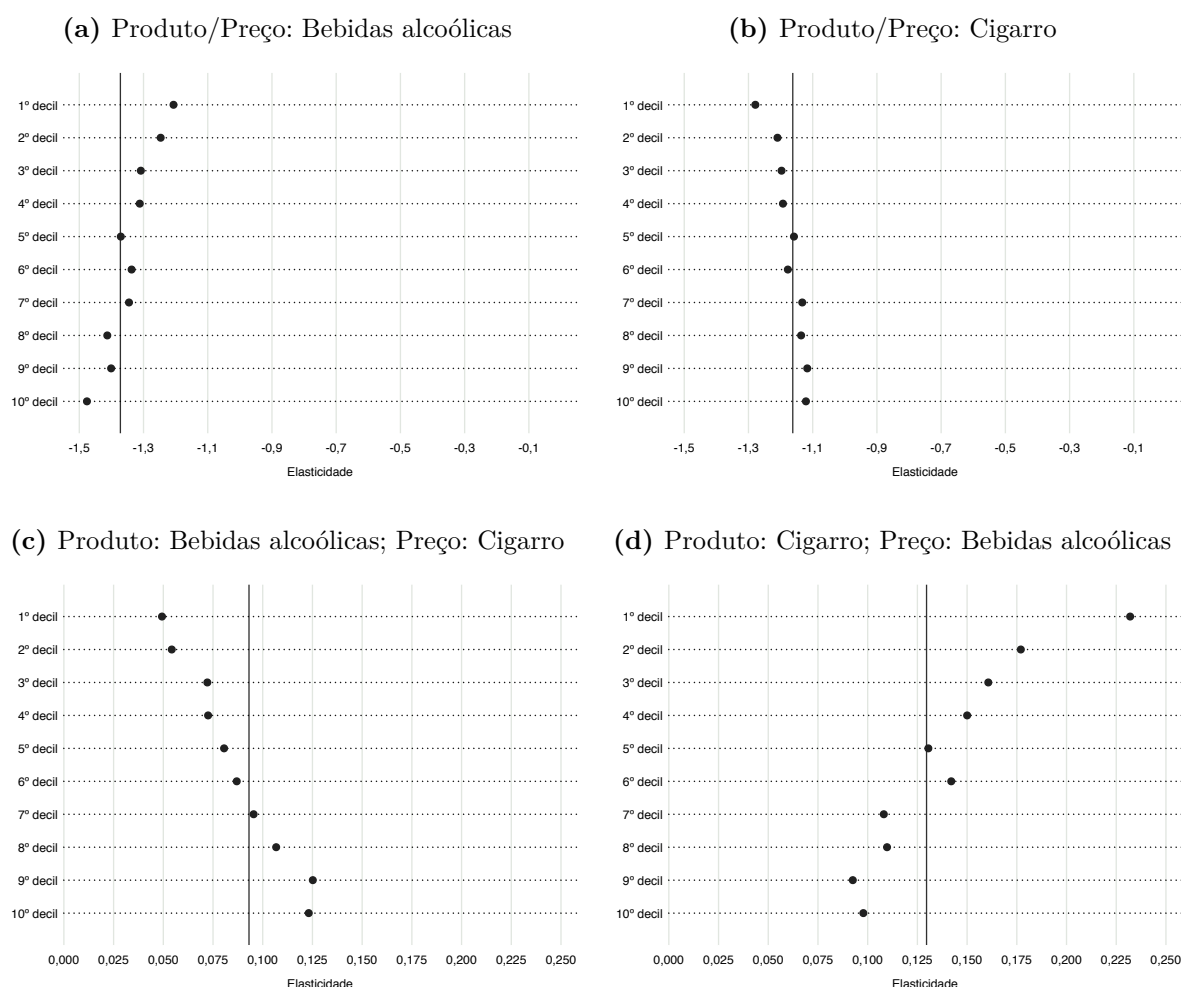
No que concerne à relação de complementariedade e substitutibilidade entre as diferentes categorias de itens as alterações nos preços do cigarro e álcool (demandas não-compensada e compensada), nota-se que todos os produtos são substitutos às bebidas alcoólicas ($E_{i7}^u > 0$), enquanto que para o cigarro metade são complementares ($E_{i8}^u < 0$) e a outra parte substitutos ($E_{i8}^u > 0$). Desta maneira, um choque nos preços das bebidas alcoólicas resultaria em uma maior demanda por outros produtos considerados, incluindo o cigarro que detém a segunda maior elasticidade preço-cruzada com a bebida alcoólica. Já para uma variação no preço do cigarro, a tendência central demonstra que parte dos efeitos cruzados gerariam, muito embora ínfimas, em uma redução na demanda por itens alimentares.

As Figuras 2.4c e 2.4d mostram os diferentes comportamentos das elasticidades-preço cruzadas para bebidas alcoólicas e cigarro por decil de renda domiciliar *per capita*, exibindo uma tendência invertida entre a magnitude da elasticidade cruzada e o decil de renda dependendo da direção dos efeitos. Quanto maior o nível de renda, maior é a sensibilidade da demanda por álcool à mudanças nos preços do cigarro, ao passo que existe uma inversão quando se avalia respostas da demanda por cigarro em relação às mudanças

¹⁶ Em Decker & Schwartz (2000), o efeito entre o preço do álcool e a quantidade demandada por cigarro é de complementariedade, enquanto que entre o preço do cigarro e a demanda por bebidas alcoólicas é de substitutibilidade.

nos preços das bebidas alcoólicas. Em termos de políticas tributárias, por exemplo, tais informações podem ser importantes para calibrar reajustes tarifários, ponderando aspectos equitativos do sistema, e identificar a repercussão do ajustamento da demanda pós-mudanças tarifárias sobre o bem-estar dos agentes. Destarte, o cigarro e as bebidas alcoólicas podem se reforçar em termos de consumo, visto que para as famílias que apresentam despesas com esses dois itens em simultâneo, tais itens ocupam posição de destaque no orçamento familiar. Contudo, em relação aos preços, estes dois grupos de produtos são substitutos para todos os níveis de renda domiciliar *per capita*¹⁷.

Figura 2.4 – Elasticidades-preço das demandas Marshallianas para bebidas alcoólicas e cigarro (ambos sentidos) por decil da renda domiciliar *per capita*



Fonte: Elaboração própria a partir dos microdados da POF 2008-2009 e da PNAD 2008.

Nota: A linha vertical representa a elasticidade-preço cruzada total.

¹⁷ Não existem evidências, por exemplo, que a proibição do uso do cigarro em locais fechados no país em 1996 (que aumentou o seu custo de oportunidade) implicou em redução no consumo de álcool e que mais, recentemente, a Lei Seca (que também amplia os custos do consumo de bebidas alcoólicas para os motoristas) induziu negativamente a demanda por cigarro. O trabalho de [Koxal & Wohlgenant \(2013\)](#) reforça esta tese de substitutibilidade dos dois produtos com base em dados de consumo para os EUA.

As elasticidades-preço da demanda compensada, que ponderam a influência do efeito renda, exibem as seguintes alterações quando confrontadas com a demanda não-compensada: as categorias de produtos com baixo grau de complementariedade passam a ser denotadas como substitutas¹⁸ (vide, por exemplo, a medida para bebidas não-alcoólicas na linha do cigarro para as demandas Marshallianas e Hicksianas); os que já possuíam a classificação de produtos substitutos têm suas magnitudes ampliadas – como a relação de substituição entre cigarro e álcool; e os que continuam como complementares, apesar do efeito renda, têm uma redução absoluta no coeficiente. Nesta pesquisa, as elasticidades Hicksianas, tendo por base a Equação 2.12, são usadas para o cálculo da variação compensatória, medida de bem-estar usada para as simulações de choques nos preços.

2.5.3 Mudanças nos preços e variação do bem-estar

Esta subseção exhibe os impactos no bem-estar em função dos choques de preços usando a VC com e sem efeito substituição, supondo uma maior tarifa de imposto que resulta em um aumento de 10% no preço ao consumidor para as bebidas alcoólicas e cigarro. Este ambiente de variação de preços está em consonância com as diretrizes atuais da política tributária brasileira, que preveem reajustes nas alíquotas do Imposto sobre Produtos Industrializados (IPI), Programa de Integração Social (PIS)/Programa de Formação do Patrimônio do Servidor Público (PASEP) e Contribuição para Financiamento da Seguridade Social (COFINS) para essas categorias de produtos¹⁹. A Tabela 2.7 exhibe a variação compensatória sem e com ajustamento de demanda, medida em termos de percentuais de dispêndio, para os consumidores de bebidas alcoólicas e cigarro por decil de renda para cada grande região do país.

¹⁸ No caso da elasticidade-preço cruzada do cigarro em relação às bebidas alcoólicas, feito apenas para demanda Marshalliana no trabalho de Pintos-Payeras (2009), esse coeficiente poderia ser positivo pela demanda compensada, dada à baixa relação de complementariedade ($E_{87}^u = -0,016$) e um efeito-renda de 0,046 (valor observado no presente estudo para este efeito).

¹⁹ Esse valor assumido nesta pesquisa é para fins de avaliação de políticas tributárias ou choques de oferta, não reproduzindo o valor expresso nas leis. A título de exemplo, a Portaria do Ministério da Fazenda N° 181 de 31/03/2014 presume uma alta em torno de 6% na tarifa para a cerveja quando comparado com as alíquotas vigentes no decreto N° 7.820 de 2012.

Tabela 2.7 – Variação compensatória com e sem efeito substituição (ES) para os consumidores de bebidas alcoólicas e cigarro, conforme decil de renda e macrorregiões (Perda de bem-estar) – em %

Decil de renda	NO	NE	SE	SU	CO	Total	NO	NE	SE	SU	CO	Total
	<i>Álcool – Sem ES</i>						<i>Álcool – Com ES</i>					
1º	5,57	4,89	4,18	4,50	5,89	4,73	5,52	4,83	4,10	4,35	5,82	4,68
2º	7,05	4,47	4,96	4,92	5,97	4,83	6,98	4,38	4,90	4,86	5,90	4,77
3º	5,80	5,07	4,50	3,06	4,96	4,61	5,75	4,97	4,41	3,01	4,89	4,56
4º	6,64	5,34	4,42	3,65	3,80	4,70	6,56	5,26	4,36	3,58	3,73	4,60
5º	7,32	5,07	4,17	2,03	4,93	4,40	7,21	5,00	4,10	2,00	4,86	4,35
6º	4,90	5,49	5,21	3,61	5,04	4,70	4,85	5,43	5,14	3,55	4,95	4,63
7º	8,48	6,13	5,80	3,65	6,04	5,03	8,40	6,04	5,73	3,58	5,96	4,95
8º	6,76	5,65	4,44	3,93	5,03	4,47	6,69	5,56	4,38	3,84	4,96	4,38
9º	5,06	4,80	5,11	3,48	6,46	4,53	5,00	4,73	5,03	3,42	6,38	4,44
10º	4,25	5,51	4,69	4,45	5,50	4,79	4,18	5,43	4,60	4,37	5,39	4,70
Total	6,12	5,21	4,67	3,70	5,27	4,68	6,05	5,14	4,60	3,63	5,19	4,61
	<i>Cigarro – Sem ES</i>						<i>Cigarro – Com ES</i>					
1º	1,34	1,71	3,50	3,64	3,31	1,97	1,30	1,66	3,42	3,55	3,24	1,93
2º	1,81	1,64	5,13	3,04	5,85	2,59	1,76	1,60	5,03	2,96	5,79	2,52
3º	1,72	1,77	4,61	3,87	3,71	2,85	1,66	1,72	4,53	3,78	3,62	2,79
4º	1,98	2,33	4,48	2,54	3,37	3,11	1,94	2,27	4,38	2,48	3,30	3,06
5º	1,75	2,37	4,48	3,59	3,79	3,33	1,71	2,31	4,39	3,49	3,70	3,25
6º	2,51	2,22	3,90	3,60	3,57	3,38	2,45	2,16	3,83	3,52	3,50	3,30
7º	2,31	2,90	5,35	3,38	4,22	4,58	2,27	2,84	5,27	3,30	4,13	4,49
8º	2,01	3,28	4,84	3,88	3,64	4,36	1,96	3,22	4,75	3,80	3,57	4,28
9º	2,65	3,31	6,29	4,17	4,17	5,00	2,61	3,26	6,19	4,09	4,10	4,91
10º	2,95	3,76	5,08	4,39	5,67	4,79	2,91	3,70	5,01	4,31	5,61	4,73
Total	1,87	2,11	4,91	3,59	4,09	3,52	1,82	2,05	4,83	3,51	4,00	3,45

Fonte: Elaboração própria a partir dos microdados da POF 2008-2009 e da PNAD 2008. Valores calculados para mediana.

NO = Norte; NE = Nordeste; SE = Sudeste; SU = Sul; CO = Centro-Oeste

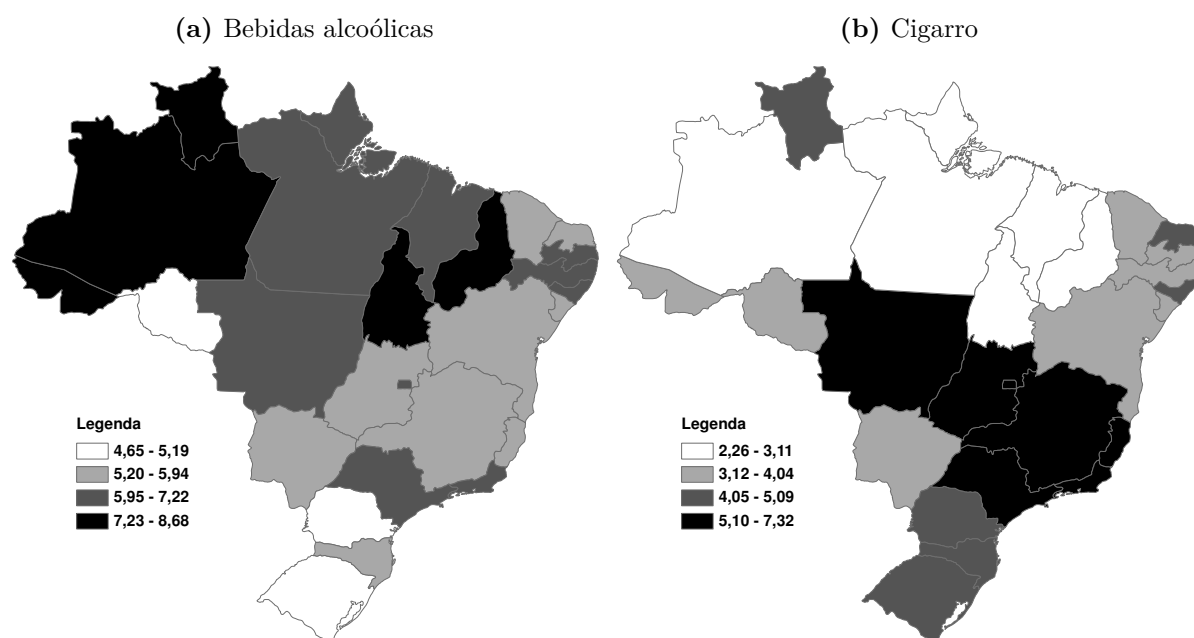
Conforme a Tabela 2.7, existem perdas para os consumidores em todos os decis de renda e agrupamentos regionais, com mais intensidade para as bebidas alcoólicas. No geral, um choque de 10% no preço das bebidas alcoólicas requer uma compensação de renda de 4,61% do dispêndio total para manutenção do nível de utilidade inicial (pré-mudança de preços), já para o cigarro a taxa de compensação é de 3,45% do dispêndio total presumindo uma alta de 10% no preço do produto. Em ambos os cenários, o ajustamento da demanda para a totalidade dos casos é de apenas 0,07 pontos percentuais (p.p.), evidenciando um baixo saldo das relações de complementariedade e substitutibilidade para essas categorias em relação aos demais itens alimentares do vetor de consumo.

Outro fator de destaque nos resultados de simulação, diz respeito ao caráter equitativo de políticas tarifárias sobre o preço do cigarro, visto que as maiores taxas de perdas ocorrem, com pequenas oscilações, para os níveis mais altos de renda e em direção das regiões mais desenvolvidas do país em termos socioeconômicos. Enquanto que a taxa de perda para uma família no primeiro decil de renda da região Nordeste é de 1,66% da despesa total, para uma família do décimo decil na mesma região este valor é de 3,70%. Em termos regionais, uma família do Sudeste na base da distribuição de renda *per capita* exibe uma perda de 3,55%, um índice 114% superior ao verificado para uma família nordestina com mesmo perfil de renda.

Ao detalhar a variação compensatória por unidade federativa, depreende-se, através

da Figura 2.5, que alterações nos preços desses produtos têm efeitos distintos entre os estados brasileiros. Contudo, de toda forma, os indicativos em nível macrorregional expressos na Tabela 2.7 são ratificados. No cenário de mudanças de tarifas das bebidas alcoólicas, as famílias com maior taxa relativa de perda, pertencentes a última classe do intervalo, estão localizadas nos Estados do Tocantins (8,68%), Acre (8,44%), Roraima (8,22%), Amazonas (7,88%) e Piauí (7,56%). Já para o cigarro, as unidades que auferem as maiores taxas de variação compensatória são Distrito Federal (7,32%), Rio de Janeiro (6,66%), São Paulo (5,94%), Espírito Santo (5,71%), Minas Gerais (5,55%), Mato Grosso (5,49%) e Goiás (5,44%). Pela distribuição desses resultados, fica claro que as perdas decorrentes de choques nos preços das bebidas alcoólicas são mais localizadas no Norte e Nordeste, à medida que a VC relativa ocorre mais fortemente nos estados dos eixos Centro-Sul.

Figura 2.5 – Variação compensatória com ajustamento de demanda para os consumidores de bebidas alcoólicas e cigarro, conforme unidade federativa (Perda de bem-estar) – em %



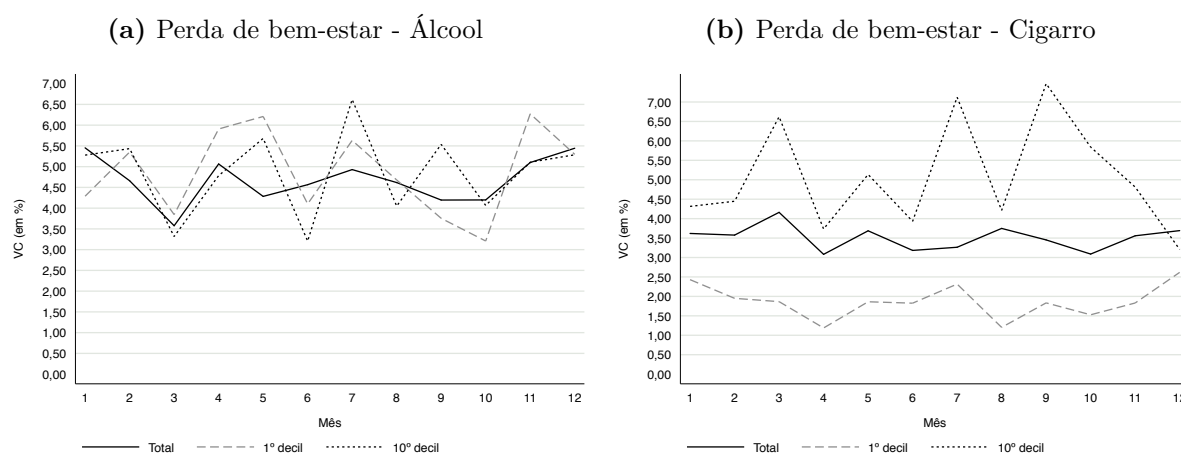
Fonte: Elaboração própria a partir dos microdados da POF 2008-2009 e da PNAD 2008.

Nota: Intervalo de classes construído pelo método de otimização de Jenks (quebrais naturais) para os valores da mediana.

A Figura 2.6 ilustra o comportamento da VC para famílias pobres (primeiro decil de renda), ricas (último decil de renda) e total ao longo dos meses de aplicação da POF para diferentes amostras de famílias. A intenção é verificar o comportamento desses indicadores em diferentes períodos de tempo, pois é possível visualizar até que ponto uma elevação de imposto para bebidas alcoólicas em épocas festivas, tais como carnaval e festas juninas, podem ter diferentes repercussões ou não para os agentes.

Verifica-se que, para as bebidas alcoólicas, as taxas de perdas entre ricos e pobres

Figura 2.6 – Variação compensatória com ajustamento de demanda no Brasil por famílias pobres, ricas e total ao longo dos meses de aplicação da POF – em %



Fonte: Elaboração própria a partir dos microdados da POF 2008-2009 e da PNAD 2008. Valores calculados para mediana, onde os meses representam o período de referência da aplicação da POF no domicílio entrevistado.

possuem praticamente as mesmas direções ao longo do tempo, contabilizando o fato de que na maior parte do período, as famílias mais pobres apresentam perdas relativas ligeiramente maiores do que as mais ricas. Já para o cigarro, o comportamento revela que ao longo dos 12 meses as famílias ricas têm taxas de perdas acima da média e bem acima dos mais pobres, onde as maiores diferenças entre os dois estratos ocorrem nos meses de setembro (5,66 p.p.), julho (4,81 p.p.) e março (4,75 p.p.). A princípio, as Figuras 2.6a e 2.6b apenas ratificam que a taxa de variação compensatória para as bebidas não possuem muita distinção entre ricos e pobres, e para o cigarro a taxa necessária para as famílias com maiores níveis de renda restaurar a utilidade pré-mudanças nos preços é superior aos das unidades mais pobres, independentemente do período da pesquisa. Ao analisar o comportamento das taxas de variação compensatória de renda para o total, não constata-se nenhuma tendência sazonal nas taxas de perdas nas duas figuras.

Como os consumos de cigarro e bebidas alcoólicas associam-se com uma série de externalidades negativas para a sociedade, as políticas tarifárias restritivas podem ser utilizadas para desestimular a sua procura, bem como gerar receitas para custear os serviços públicos de saúde, por exemplo. O somatório das perdas das famílias em respostas ao choque de 10% nos preços do álcool e cigarro totalizam apenas 3,1% dos gastos estimados com as DCNT no período pelo Sistema Único de Saúde (SUS), com as despesas de procedimentos ambulatoriais e internações²⁰. Adverte-se que os custos da sociedade com as externalidades do uso de cigarro e bebidas alcoólicas incluem, além das despesas do SUS, absenteísmo, aposentadorias precoces, perda de produtividade etc. (BRASIL, 2005), bem

²⁰ Estimativa de despesa calculada com base em Brasil (2005), que estimam esses custos para o ano de 2002 e a presente pesquisa usa o IGP-DI para corrigir esse valor para janeiro de 2009.

como um elevado número de mortes relacionados direta e indiretamente com a utilização desses produtos ([WHO - World Health Organization, 2011](#)).

2.6 Conclusões

Através do modelo QUAIDS com ajustes para o consumo censurado e endogeneidade dos gastos totais, observa-se que a demanda pelas nove categorias de produtos mantém relações com os preços, renda e fatores conjunturais e locais. Tais fatos ilustram a necessidade de se compreender melhor a demanda por produtos que estão na lista dos fatores de risco modificáveis à saúde que mais causam mortes no Brasil, de modo a identificar como alterações motivadas por políticas públicas podem repercutir sobre os seus consumidores.

A elasticidade-dispêndio da demanda por cigarro e bebidas alcoólicas é positiva, mas menor do que um para o primeiro produto. Para a bebida alcoólica, um choque positivo na renda propicia uma ampliação proporcionalmente maior em sua procura para qualquer estrato de renda domiciliar *per capita* e região de localização da família. Além do mais, dentre as nove categorias de produtos, famílias pobres que fazem parte do grupo de interesse das políticas sociais de transferência de renda, têm uma das maiores elasticidade-dispêndio em relação à procura por bebidas com teor alcoólico, sugerindo que o crescimento na demanda por álcool responde mais fortemente a incrementos na renda do que o consumo de itens alimentares (como cereais, massas, frutas e verduras, carnes e leite) e bebidas não-alcoólicas.

Pela matriz de elasticidade-preço das demandas compensadas e não-compensadas, o cigarro e as bebidas alcoólicas possuem uma relação de substitutibilidade em termos de preços cruzados em ambos os sentidos. Essa informação é importante para identificar que modificações nos preços de um desses produtos não originam redução de demanda do outro. Alicerçados nesses indicadores, a simulação de um aumento tarifário nos produtos mencionados sobre o bem-estar das famílias assinalam a existência de um baixo ajustamento de demanda para o vetor de consumo avaliado. Deste modo, a taxa de compensação de renda estimada é maior para as famílias e regiões mais ricas ao longo dos 12 meses do período de referência da POF, enquanto que para a demanda por bebidas alcoólicas a taxa de perda de bem-estar não difere muito entre os diferentes estratos de rendimento domiciliar.

Não obstante às perdas na utilidade para essas famílias resultantes de políticas tarifárias mais restritivas, é válido salientar que o somatório de todos os valores de renda requeridos para restaurar o nível de utilidade pré-mudança de preços representa apenas uma parcela ínfima, por exemplo, dos custos diretos do SUS com procedimentos ambulatoriais e internações motivadas por DCNT. Como elevações de alíquotas tarifárias apresentam

certos limites práticos, dado aspectos como sonegação fiscal, contrabando e substitutos ilícitos, pontua-se que ações voltadas para o desestímulo do consumo do cigarro e de bebidas alcoólicas devam ser mais amplas do que políticas restritivas sobre os preços destes itens, tais como a adoção de políticas não-tarifárias voltadas para as crianças e jovens a fim de minimizar o consumo no médio e longo prazo.

3 Efeitos da exposição aos fatores de risco comportamentais à saúde sobre o atraso escolar no Brasil

3.1 Introdução

Devido à importância da educação no processo de formação do capital humano, existem pesquisas específicas para a compreensão dos fatores associados ao desempenho escolar dos indivíduos. Desde meados do século XX, em especial a partir do relatório de [Coleman *et al.* \(1966\)](#), que essa temática vem sendo objeto de interesse crescente para a área acadêmica e para os *policy makers*. Para o Brasil, a grande parte das pesquisas sobre os determinantes de indicadores educacionais examinam o papel das condições socioeconômicas do discente, das características dos insumos escolares e dos programas sociais ([GOMES-NETO; HANUSHEK, 1994](#); [BARROS *et al.*, 2001](#); [SOARES, 2003](#); [MACHADO, 2008](#); [RIANI; RIOS-NETO, 2008](#); [SAMPAIO *et al.*, 2011](#); [ALMEIDA, 2014](#)).

Adicionalmente, a literatura internacional examina as consequências no desempenho na escola em razão da exposição dos jovens aos fatores de risco à saúde ([JAMISON, 1986](#); [GLEWWE; JACOBY, 1995](#); [YAMADA; KENDIX; YAMADA, 1996](#); [CROSNOE; MULLER, 2004](#); [KAESTNER; GROSSMAN, 2009](#); [KAESTNER; GROSSMAN; YARNOFF, 2011](#); [CARRELL; HOEKSTRA; WEST, 2011](#); [PONZO, 2013](#); [LINDO; SWENSEN; WADDELL, 2013](#)). As condições de saúde dos indivíduos, decorrentes de preferências de risco, possuem impactos diretos no funcionamento individual, com correspondência em fatores: físicos, como limitações funcionais, problemas cardiovasculares e respiratórios, maiores incidências de diabetes e outros agravos à saúde; não-físicos, como baixa concentração, problemas psicológicos, maior absenteísmo e menor participação em sala de aula ([CHALOUPIKA; WARNER, 1999](#); [KENKEL; WANG, 1999](#); [CAWLEY; RUHM, 2011](#)). Por decorrência, seguindo a linha desses autores, a criança exposta a hábitos não saudáveis tende a apresentar menor rendimento na escola.

A Pesquisa Nacional de Saúde do Escolar (PENSE) de 2012, para alunos do último ano do ensino fundamental no Brasil, mostra que 67% dos estudantes já consumiram bebidas alcoólicas, 20% foram expostos ao uso do cigarro e cerca de 40% dos discentes passam no mínimo 5 horas por dia (exclusive feriados e fins de semana) sentados – assistindo televisão, ao computador ou jogando videogame. Esses números revelam que uma quantidade elevada de adolescentes no país adotou ou adota posturas de risco à saúde em idade escolar, sendo relevante examinar as implicações dessas exposições.

É possível encontrar trabalhos aplicados para dados nacionais, como [Gomes-Neto *et al.* \(1997\)](#) e [Machado \(2008\)](#), que estimam o impacto das condições de saúde dos indivíduos, através de indicadores antropométricos, sobre os resultados escolares. Contudo, ainda são escassos estudos que associam hábitos não saudáveis das crianças e rendimento escolar. A preocupação com os chamados fatores de risco modificáveis faz parte inclusive de um recente plano de ações estratégicas, lançado pelo governo brasileiro, para o enfrentamento das Doenças Crônicas Não Transmissíveis (DCNT), com metas de redução do crescimento das DCNT no período de 2011 a 2022 ([BRASIL, 2011](#)).

Assim, este trabalho pretende avançar nas discussões dos determinantes do desempenho educacional no Brasil ao avaliar o efeito de comportamentos não saudáveis, representados pela exposição ao cigarro, à bebida alcoólica e aos problemas de sedentarismo e má alimentação (captado pelo excesso de peso), sobre a distorção idade-série de estudantes no 9º ano do ensino fundamental. Esta abordagem possibilita a integração das literaturas de teoria do capital humano, economia da educação e da saúde, uma vez que as experiências e acúmulos de conhecimentos nessa fase da vida servem de base para as trajetórias econômicas, sociais e de saúde dos indivíduos no longo prazo ([CONTI; HECKMAN; URZUA, 2010](#)). O foco na avaliação dos comportamentos em destaque sobre os resultados na escola está baseado em [Cawley & Ruhm \(2011\)](#), que assinalam os fortes efeitos diretos de fumar, beber e sobrepeso na própria saúde.

Este capítulo está dividido em seis partes, o que inclui esta introdução. A Seção 3.2 apresenta os principais resultados da literatura baseados na função de produção educacional, enfatizando o papel das condições de saúde dos alunos. A Seção 3.3 e a Seção 3.4 tratam, respectivamente, da estratégia empírica e das informações sobre a base de dados e tratamento das variáveis. Por fim, a Seção 3.5 e a Seção 3.6 mostram os principais resultados e conclusões.

3.2 Função de produção educacional e o papel das condições de saúde

A partir do relatório de [Coleman *et al.* \(1966\)](#), baseado em uma análise detalhada sobre a igualdade de oportunidades educacionais nas escolas públicas norte-americanas no ano de 1965, existe uma preocupação crescente na literatura especializada em identificar os determinantes do resultado escolar, seja em países desenvolvidos ou em desenvolvimento. A função de produção educacional que baliza esses trabalhos é definida na Equação 3.1

$$A_i = f(S_i, Q_i, C_i, H_i, I_i), \quad (3.1)$$

onde: A_i é o desempenho escolar do i -ésimo indivíduo; S é uma variável escalar que mede os anos de estudo; \mathbf{Q} representa o vetor de atributos da escola e dos professores; \mathbf{C} é o vetor de variáveis individuais do aluno; \mathbf{H} é o vetor de características da família; \mathbf{I} é o vetor de insumos escolares que estão sob o controle dos pais, tais como a compra de livros e outras necessidades escolares.

Através da função de produção educacional, estudos como Hanushek (1970, 1971, 1986), Eide & Showalter (1998), Rangvid (2007) e Glewwe *et al.* (2011), em nível internacional, e Barros *et al.* (2001), Albernaz, Ferreira & Franco (2002), Machado *et al.* (2008), Riani & Rios-Neto (2008) e Almeida (2014), em nível nacional, enfatizam o papel do *background* familiar no processo de geração do bom desempenho acadêmico do aluno. Glewwe *et al.* (2011) fazem uma revisão da literatura nas áreas de educação e economia de 1990 a 2010 nos países em desenvolvimento, tendo como meta investigar as características da escola e do professor no resultado escolar a partir de uma seleção de 79 publicações. Através dessa análise, os autores concluem que as características da escola e do professor possuem pequenos efeitos ou insignificância estatística, levando em conta a qualidade dos modelos estatísticos. Independentemente do agrupamento econômico do país, existe certo consenso acerca da importância do perfil do aluno e da família na produção da educação.

No campo da literatura que sublinha o papel dos atributos de saúde das crianças e adolescentes, que compõem o vetor de características individuais, nos indicadores de realização educacional é recorrente a identificação de uma associação positiva entre boas condições de saúde e medidas de *performance* educacional (EIDE; SHOWALTER, 2011). Os trabalhos apontam os efeitos do status nutricional, como Jamison (1986), Moock & Leslie (1986), Glewwe & Jacoby (1995) e Gomes-Neto *et al.* (1997), de fatores comportamentais, como Pirie, Murray & Luepker (1988), Ellickson, Tucker & Klein (2001), Carrell, Hoekstra & West (2011) e Lindo, Swensen & Waddell (2013), e outras medidas de saúde, como Eide, Showalter & Goldhaber (2010), Rees & Sabia (2011) e Rees & Sabia (2014), sobre o nível de proficiência, atraso escolar, conclusão dos estudos e outros indicadores educacionais.

Jamison (1986) estima a relação entre a saúde das crianças chinesas – através de razões normalizadas de altura-idade, peso-idade e peso-altura – e o desempenho na escola para uma amostra com 3.000 estudantes do ensino fundamental em cinco regiões da China. Os seus achados indicam que o baixo status nutricional, mensurado pela altura-idade, atua negativamente na *performance* estudantil independentemente da região. Nessa mesma direção, o trabalho de Moock & Leslie (1986) é desenvolvido para uma amostra final de 350 crianças entre 5 e 11 anos de idade de escolas primárias da região de Terai no Nepal. As estimativas feitas por mínimos quadrados ordinários (MQO) e por um modelo probit pontuam que a desnutrição induz baixos indicadores escolares (matrícula e promoção na escola). Os autores ressaltam que esforços para melhorar o estado nutricional da criança pode implicar, além dos benefícios para a saúde, em maiores realizações educacionais.

Partindo da hipótese que o atraso escolar são respostas racionais a desnutrição na primeira infância, [Glewwe & Jacoby \(1995\)](#) usam dados do *Ghana Living Standards Survey* (GLSS) 1988-1989 e um modelo probit ordinal com diferentes alternativas de especificação para uma amostra final de 1.757 crianças entre 6 e 15 anos de idade. As estimativas assinalam que a desnutrição na primeira infância causa o ingresso tardio na escola, implicando a pertinência de políticas de intervenção nutricional na primeira infância com vistas na melhoria da saúde e aumento futuro da riqueza desses indivíduos.

A partir dos dados do *National Longitudinal Study of Adolescent Health* (NLSAH) de 2004 e regressão multinível, [Crosnoe & Muller \(2004\)](#) verificam que a obesidade, por ser uma condição de saúde estigmatizada na população norte-americana, relaciona-se de forma negativa com o desempenho acadêmico dos alunos do ensino secundário. Esse impacto é ainda maior nas escolas que, por exemplo, possuem altas taxas de alunos participantes de atividades esportivas. Os autores argumentam que o baixo desempenho na escola, motivado por questões de obesidade, tem fortes influências de longo prazo nas relações interpessoais. Utilizando a pesquisa *Icelandic Study* para estudantes com 14 e 15 anos de idade de escolas secundárias da Islândia, [Sigfusdottir, Kristjansson & Allegrante \(2007\)](#) revelam que comportamentos saudáveis (índice de massa corporal, dieta nutricional, prática de atividades físicas e fatores de saúde mental) são importantes para melhoria do desempenho acadêmico.

Pelas estimações do modelo empírico de [Kaestner & Grossman \(2009\)](#), baseado nos dados do *National Longitudinal Survey of Youth* (NLSY) para estudantes entre 14 e 18 anos de idade e nos indicadores de progressão e abandono escolar, uma pior realização acadêmica não pode ser atribuída ao excesso de peso na adolescência. Em contraste, [Florin, Shults & Stettler \(2011\)](#) realçam a importância da percepção do excesso de peso na compreensão dos resultados acadêmicos dos adolescentes, pois essa relação negativa entre o desempenho escolar e a saúde do jovem tem expressivas consequências para o nível de bem-estar do indivíduo quando adulto.

Afora estudos que avaliam as relações nutricionais (desnutrição ou, no outro extremo, excesso de peso), existem pesquisas que abordam os efeitos de aspectos comportamentais e outras medidas de saúde no desempenho escolar. [Rees & Sabia \(2011\)](#) examinam o efeito da enxaqueca sobre a média de notas no ensino médio, probabilidade de concluir o ensino médio e probabilidade de cursar uma universidade. Usando dados do NLSAH em conjunto com as abordagens MQO e *propensity score matching* (PSM), as estimativas encontradas sugerem que a enxaqueca está negativamente relacionada com todos os três indicadores educacionais.

Com base nas pesquisas *Progress in International Reading Literacy Study* (2006-PIRLS) e *Trends in International Mathematics and Science Study* (2007-TIMSS) e em técnicas paramétricas e não-paramétricas (PSM), [Ponzo \(2013\)](#) constata uma relação

negativa entre a vítima de *bullying* na escola – que reflete se a criança foi assediada, roubada ou ferida por outro estudante – e a nota auferida em matemática, leitura e ciências. Em igual direção, [Rees & Sabia \(2014\)](#) verificam que o problema de gagueira associa-se também de forma negativa com os resultados escolares, considerando os dados do NLSAH e um conjunto de modelos econométricos, dentre eles o PSM.

Segundo [Lye & Hirschberg \(2010\)](#) e [Cawley & Ruhm \(2011\)](#), os efeitos do consumo de álcool, cigarro e outros tipos de drogas sobre os resultados educacionais têm sido frequentemente documentados na literatura internacional, especialmente as investigações sobre os impactos do uso de bebidas. Por exemplo, [Koch & Ribar \(2001\)](#), a partir do painel de dados do NLSY 1979-1990 para pares de irmãos do mesmo sexo, encontram que as consequências do uso de bebidas alcoólicas na juventude sobre a educação é sensível as diferentes especificações econométricas. Eles destacam que o consumo abusivo de álcool diminui os indicadores de escolaridade, mas que o uso moderado não tem um efeito expressivo no desempenho escolar. Outrossim, [Chatterji \(2006\)](#) sugere que o uso de álcool no ensino médio está associado com reduções no nível de escolaridade dos indivíduos. Nesse trabalho, as estimativas são calculadas através dos dados do *National Education Longitudinal Study* (NELS) e de modelos probit.

Ao verificar os resultados de [Renna \(2007\)](#), que utiliza os dados do NLSY de 1979, tem-se que o consumo excessivo de álcool diminui a probabilidade do indivíduo concluir o ensino médio (até 19 anos de idade) em até 5,2% para as mulheres e 14,5% para os homens. [Renna \(2008\)](#) amplia a investigação do trabalho anterior, ao considerar as consequências da utilização de bebidas com teor alcoólico na adolescência sobre um conjunto maior de *outcomes* educacionais, incluindo além da conclusão do ensino médio, indicadores como abandono escolar e conclusão na idade prevista. Através dos dados do NLSY de 1979 e de diferentes modelos estatísticos, como probit ordenado, logit multinomial e probit em dois estágios, o autor confirma a hipótese dos efeitos adversos da bebida alcoólica na realização educacional dos discentes, dada a redução na probabilidade destes alunos alcançarem um maior nível de escolaridade.

Por sua vez, [Balsa, Giuliano & French \(2011\)](#) mostram que o aumento no consumo de álcool resulta em pequenas reduções na qualidade da aprendizagem para estudantes do sexo masculino, enquanto que para as alunas os efeitos não são significativos estatisticamente. Os autores usam dados do NLSAH e modelos de efeitos fixos para analisar o papel do uso de álcool sobre indicadores educacionais quantitativos (dias de aulas perdidos) e qualitativos (dificuldades com concentração na escola, convivência com os professores ou completar a lição de casa). No caso dos meninos, a redução na performance educacional ocorre em particular no aumento de faltas às aulas e nas dificuldades de realização das tarefas escolares.

Com o objetivo de avaliar o efeito do consumo de álcool sobre o desempenho escolar,

Carrell, Hoekstra & West (2011) investigam o papel da lei da idade mínima para beber nos Estados Unidos. Os autores fazem uso de regressão descontínua e os microdados de discentes do *Air Force Academy* (USAFA) no período de 2000 a 2006, encontrando que a utilização de bebida alcoólica provoca reduções significativas no desempenho acadêmico, especialmente para os alunos com melhor *performance*. Nessa mesma linha, usando o método de regressão descontínua e os dados de estudantes – que ingressaram no curso com 18 ou 19 anos de idade – da Universidade de Oregon entre 1999 e 2007, Lindo, Swensen & Waddell (2013) examinam o efeito do acesso legal às bebidas alcoólicas no desempenho acadêmico. Os resultados centrais da pesquisa indicam que as notas dos discentes são inferiores aos níveis esperados em cerca de 0,03 desvios-padrão, quando se mede o efeito do acesso legal às bebidas. A consequência dessa lei é maior para as mulheres, homens com baixos níveis de habilidades e jovens mais desfavorecidos economicamente.

Outros resultados sobre o uso de substâncias que causam dependência – como cigarro, maconha e *ecstasy* – na realização acadêmica podem ser verificados também na literatura internacional. Através de uma pesquisa de acompanhamento inicial de 7.124 adolescentes de quatro distritos da área metropolitana de Minneapolis-Saint Paul nos Estados Unidos entre os anos de 1979 e 1985, Pirie, Murray & Luepker (1988) identificam que os alunos que abandonam os estudos possuem elevadas taxas de uso frequente do tabaco. Enquanto que a taxa de prevalência de fumantes em idade escolar que estavam matriculados em 1985 era em torno de 16,6% entre os meninos e de 22,5% entre as meninas, essa taxa era superior a 77% entre os indivíduos que haviam abandonado os estudos independentemente do gênero.

Em Yamada, Kendix & Yamada (1996), a análise dos hábitos prejudiciais incluem além do consumo de bebidas alcoólicas, o uso de maconha por parte de alunos no ensino médio. Considerando os dados do NLSY e o modelo probit, os autores concluem que existem significantes efeitos adversos da adoção desses comportamentos prejudiciais à saúde nos resultados educacionais, haja vista a redução na probabilidade de conclusão do ensino médio. Para Yamada, Kendix & Yamada (1996), a utilização abusiva de substâncias ilícitas, como a maconha, implica em grandes perdas na aquisição de capital humano, sendo altamente recomendado programas preventivos que desencorajem a adoção de hábitos não saudáveis.

Baseado nos dados longitudinais de estudantes da Califórnia e de Oregon e na regressão logística, Ellickson, Tucker & Klein (2001) verificam que o uso prematuro do cigarro por parte dos discentes entre a 7ª série e a 12ª série da educação básica gera maiores dificuldades acadêmicas e tendências de uso regular de substâncias nocivas e comportamentos delinquentes. No comparativo com os não-fumantes, os fumantes prematuros têm ao menos três vezes mais chances de usar de forma regular tabaco e maconha, vender drogas e experimentar gravidez e paternidade precoce, bem como cinco

vezes mais chances de abandonar os estudos.

Por meio dos dados do NELS de 1992, os achados de [Jeynes \(2002\)](#) mostram que os padrões de adolescentes acerca do consumo, especialmente, de álcool e cigarro têm impactos negativos e estatisticamente significativos sobre o desempenho dos alunos em testes padronizados de matemática, leitura, ciências e estudos sociais. No que concerne ao uso do cigarro, por exemplo, os alunos fumantes exibem uma nota no cômputo total inferior em 0,15 pontos aos dos alunos que não adotam esse tipo de hábito, tendo em vista os controles socioeconômicos e adoção de outras posturas prejudiciais à saúde. Além do mais, de acordo com o trabalho de [Martins & Alexandre \(2009\)](#) com dados do *National Survey of Drug Use and Health* (NSDUH) e *Youth Risk Behavior Survey* (YRBS) e regressão logística, os não usuários de drogas – álcool, tabaco, maconha e *ecstasy* – apresentam as melhores notas na escola quando confrontados com os usuários desses itens. Para [Martins & Alexandre \(2009\)](#), em termos gerais, o uso de *ecstasy* reduz a realização acadêmica dos estudantes em uma magnitude superior a observada pelo o uso de maconha, álcool e tabaco.

No Brasil, conforme [Machado \(2008\)](#), as investigações sobre os efeitos da saúde em resultados escolares estão em fase inicial. [Gomes-Neto et al. \(1997\)](#) introduz essa discussão, utilizando dados longitudinais do projeto EDURURAL da década de 80, para menos de 400 estudantes de áreas rurais dos estados do Piauí, Ceará e Pernambuco. Apesar dos autores demonstrarem o papel de destaque da boa nutrição na realização acadêmica e no atendimento escolar, eles reconhecem que os resultados devem ser vistos com cautela devido ao restrito desenho amostral. Nessa linha, [Machado \(2008\)](#) investiga a relação entre a saúde infanto-juvenil, medida por indicador padronizado de altura-idade, e o ingresso na escola em idade considerada legalmente inadequada. Para isso, a Pesquisa sobre Padrões de Vida (PPV) 1996-1997, realizada pelo o Instituto Brasileiro de Geografia e Estatística (IBGE) para 10 estratos geográficos de localidades das regiões Nordeste e Sudeste do país, é usada concomitantemente com um modelo de resposta binária probit para uma amostra final de 2.712 observações. Esse trabalho ratifica o papel das boas condições de saúde e nutrição sobre a *performance* escolar.

3.3 Estratégia empírica

A hipótese suscitada por esta pesquisa diz respeito ao papel de comportamentos de risco à saúde na infância e na primeira adolescência sobre o desempenho escolar no Brasil. Para testar essa hipótese, as estimações dos parâmetros de interesse obedecem três estágios: I) estimação da função de distorção idade-série para exploração inicial dos resultados e primeiras estimativas da associação dos hábitos não saudáveis com o *outcome* escolar; II) cálculo do efeito médio da exposição aos fatores de risco pelas diferentes técnicas de

pareamento, com o intuito de verificar se a direção dos resultados apontados no estágio I são mantidos; III) análise da sensibilidade das estimativas do estágio anterior através dos limites de Rosenbaum para quantificar o tamanho de possíveis interferências de variáveis não-observadas.

Para [Park & Kang \(2008\)](#) e [Jurges, Reinhold & Salm \(2011\)](#), é preciso atenção na questão da endogeneidade que envolve a medição do efeito causal das relações entre as condições de saúde e a educação, pois indivíduos com mais anos de estudo podem optar pela adoção de um estilo de vida mais saudável, condizentes com as preferências intertemporais dos agentes entre o consumo presente e futuro. Por outro lado, indivíduos mais saudáveis podem priorizar os resultados da função de produção da educação. Nessa situação, o problema de endogeneidade é melhor definido quando os agentes já apresentam maturidade para formulação das preferências intertemporais. Reforçando esse argumento, [Becker & Mulligan \(1997\)](#) sublinha que a educação tem um efeito causal sobre a saúde, sempre que a escolaridade repercute nas preferências dos agentes em um dado intervalo de tempo. Muito embora, possam existir fatores não-observados específicos da criança e de seus familiares que afetem ambos indicadores de saúde e de escola, os trabalhos que direcionam a análise da saúde sobre a educação em geral restringem a amostra para crianças e adolescentes para amenizar esse tipo de problema ([EIDE; SHOWALTER, 2011](#)).

Desse modo, a estratégia adotada para mensurar os efeitos pretendidos está baseada em um desenho amostral direcionado a alunos do ensino fundamental que foram expostos aos fatores de risco comportamentais no período antes ou equivalente a pré-adolescência (faixa etária entre 11 e 14 anos de idade). Nessa fase é suposto que a simultaneidade entre educação e condições de saúde ainda não está completamente definida, pois o status de saúde da criança precede ao resultado na escola. Portanto, não existe ou é pouco provável a relação de causalidade reversa, mas sim a direção de que problemas de saúde, seja em termos de condições físicas ou psicológicas, impactam no desempenho do aluno. Ademais, a taxa de desconto sobre o valor futuro da saúde ainda pode ser tida como indefinida, visto que a criança não dispõe de maturidade e informações suficientes para formular suas expectativas quanto à maximização intertemporal da sua utilidade, condizente com o estoque de saúde futuro. As escolhas de risco nessa fase de vida é influenciada por fatores exógenos aos alunos, como o ambiente familiar e os efeitos dos pares ([CROSNOE; MULLER, 2004](#)). Mesmo admitindo essas hipóteses, análises de sensibilidade são desenvolvidas para verificar se as características não-observadas dos alunos poderiam alterar os efeitos da exposição dos fatores de risco à saúde sobre o desempenho escolar.

Como destacado na Equação 3.1, as condições de saúde da criança correspondem a um dos inputs da produção educacional, estando presente no vetor de características individuais do aluno. Como no Brasil não existem microdados disponíveis pelos órgãos estatísticos oficiais que reúnam em conjunto informações detalhadas sobre saúde do

indivíduo e resultados na escola (nível de proficiência²¹ e taxas de rendimento escolar²²), utiliza-se a distorção idade-série como indicador educacional²³.

3.3.1 Modelo logit: efeito médio não condicionado ao contrafactual observável

O modelo empírico desta parte se baseia na função de produção educacional, assumindo que o desempenho escolar é função dos insumos relativos às características dos alunos, do *background* familiar, da escola e dos fatores locais. Nos atributos específicos dos estudantes, as estimativas das exposições aos fatores de risco comportamentais à saúde são os parâmetros de interesse. A Equação 3.2, base para os primeiros resultados desta pesquisa, está descrita a seguir:

$$A_i^* = f(\beta_0 + \beta_1 H_{1i} + \beta_2 H_{2i} + \beta_3 H_{3i} + \mathbf{X}_i \mathbf{b} + u_i), \quad (3.2)$$

onde: A_i^* é a variável latente relacionada ao desempenho escolar do i -ésimo indivíduo; H_1, H_2 e H_3 são variáveis binárias de exposição aos fatores de risco pelo o indivíduo i , respectivamente, uso de cigarro, álcool e excesso de peso; \mathbf{X}_i é um vetor que contém as demais variáveis explanatórias que representam outras características específicas do aluno (raça, gênero), da família (anos de estudos da mãe e do pai, nível socioeconômico, estrutura do núcleo familiar, tamanho da família), da escola (recursos de infraestrutura, dependência administrativa) e fatores locais (região de localização da escola).

Supondo que a variável qualitativa distorção idade-série (dis_i) segue distribuição logística, é possível determinar a probabilidade de que A_i^* seja inferior ou igual a um determinado limiar (j), isto é: $Pr(A_i^* \leq j) = G(\cdot)$. Assim, a Equação 3.3 é desenvolvida para o caso em que dis_i é enquadrado numa resposta binária, com: $dis_i = 0$, se $A_i^* \leq 0$; e $dis_i = 1$, se $A_i^* > 0$:

$$Pr(dis_i = 1) = G(\beta_0 + \beta_1 H_{1i} + \beta_2 H_{2i} + \beta_3 H_{3i} + \mathbf{X}_i \mathbf{b} + u_i), \quad (3.3)$$

sendo $G(\Theta) = 1/[1 + \exp(-\Theta)]$.

Os procedimentos adotados para o cálculo do efeito médio da exposição aos fatores de risco sobre a probabilidade de distorção idade-série nessa primeira abordagem estão

²¹ Atualmente, o Brasil dispõe de um conjunto de testes padronizados aplicados para os estudantes da educação básica, como a Provinha Brasil, a Prova Brasil e o ENEM. Contudo, nenhum desses exames apresentam dados sobre comportamentos não saudáveis dos alunos e outras medidas de saúde.

²² As taxas de rendimento escolar indicam o fluxo do aluno na escola. As informações de rendimento escolar são sumarizadas por indicadores de aprovação, repetência e abandono escolar.

²³ Esse tipo de estratégia é bastante comum na literatura, que usam medidas referentes ao ingresso tardio na escola e outras medidas de atraso escolar (GOMES-NETO; HANUSHEK, 1994; GLEWWE; JACOBY, 1995; RIOS-NETO; CESAR; RIANI, 2002; MACHADO; GONZAGA, 2007; MACHADO, 2008; RIBEIRO; CACCIAMALI, 2012).

descritos na Equação 3.4

$$EMD_k = n^{-1} \sum_{i=1}^n [G(\hat{\beta}_0 + \hat{\beta}_k + \mathbf{W}_i \hat{\mathbf{c}}) - G(\hat{\beta}_0 + \mathbf{W}_i \hat{\mathbf{c}})], \quad (3.4)$$

em que \mathbf{W}_i é o vetor de covariadas com exceção da variável binária de exposição ao fator de risco k .

Como destaca Wooldridge (2010), o cálculo do efeito parcial médio²⁴ (EPM) e do efeito parcial na média²⁵ ($EPPM$) não fazem sentido para variáveis explicativas discretas, como é a situação de H_1, H_2 e H_3 , que medem a exposição aos fatores de risco comportamentais, por isso a necessidade de estimar o efeito dessas covariadas de interesse através da Equação 3.4 que expressa a média de alteração na probabilidade dos alunos estarem com atraso escolar quando se compara à situação de não-exposição ao fator de risco H_k .

3.3.2 Modelo PSM: efeito médio condicionado ao grupo de controle observável

Esta subseção descreve brevemente a abordagem *propensity score matching* (PSM), usada para o cálculo do efeito médio baseado na identificação de um grupo de controle com similar características observáveis (\mathbf{Z}_k) dos discentes expostos. Este método proposto por Rosenbaum & Rubin (1983) resume as características de cada unidade em um único escore que viabiliza o pareamento.

As dificuldades para medir o efeito causal de uma determinada covariada diz respeito ao problema do contrafactual, dado que só é possível observar apenas um dos resultados potenciais $Y_i(1)$, devido à indisponibilidade da informação sobre o resultado do indivíduo tratado caso ele não fosse exposto, isto é, o seu resultado contrafactual é não observável, $Y_i(0)$. Como essa informação é um *missing value*, torna-se fundamental a construção do contrafactual do grupo exposto aos fatores de risco, denotado como grupo de controle – indivíduos não-expostos, mas que guardam características similares ao grupo tratado.

O primeiro estágio do PSM requer a estimação da probabilidade do aluno ser exposto ao fator de risco k , sendo necessário dispor de fatores que influenciam essa decisão, bem como covariadas associadas com a variável de resultado. Os modelos empíricos na área da economia da saúde, como Cutler & Glaeser (2005), Park & Kang (2008), Cutler & Lleras-Muney (2010), Rees & Sabia (2011), Jorges, Reinhold & Salm (2011) e Rees & Sabia (2014), mostram que a decisão do indivíduo por hábitos não saudáveis é influenciada, especialmente, pelo *background* familiar e atributos individuais e situacionais.

²⁴ $EPM_j = n^{-1} \sum_{i=1}^n [g(\hat{\beta}_0 + \hat{\beta}_1 H_{1i} + \hat{\beta}_2 H_{2i} + \hat{\beta}_3 H_{3i} + \mathbf{X}_i \hat{\mathbf{b}}) b_j]$, sendo $g(\cdot) = \exp(\cdot) / [1 + \exp(\cdot)]^2$.

²⁵ $EPPM_j = g(\hat{\beta}_0 + \hat{\beta}_1 \bar{H}_{1i} + \hat{\beta}_2 \bar{H}_{2i} + \hat{\beta}_3 \bar{H}_{3i} + \bar{\mathbf{X}}_i \hat{\mathbf{b}}) b_j$, sendo \bar{H}_{ki} e $\bar{\mathbf{X}}_i$ as médias amostrais das variáveis.

Tendo o suporte desses fatores e da função de produção educacional, a probabilidade condicional de exposição ao fator de risco k condicionado a um vetor de características observáveis é definida por

$$Pr(H_k = 1|\mathbf{Z}_k) = G(\mathbf{Z}_k\boldsymbol{\lambda} + \epsilon), \quad (3.5)$$

onde: H_k é o indicador de exposição ao fator de risco saúde k ; \mathbf{Z}_k é o vetor de características observáveis para o fator de risco k , composto pelo perfil do aluno, da família, da escola e de covariadas específicas atreladas a decisão de exposição; G representa a função de distribuição acumulada. Supondo que G segue uma distribuição de probabilidade logística, pode-se estimar $Pr(\mathbf{Z}_k)$ através de um modelo de resposta qualitativa logit.

Após a obtenção do escore de propensão $Pr(\mathbf{Z}_k)$, o efeito médio do tratamento sobre o grupo exposto (τ_k) é estimado a partir da Equação 3.6

$$\tau_k = E_{Pr(\mathbf{Z}_{ki})|H_k=1}\{E[dis_i|H_{ki} = 1, Pr(\mathbf{Z}_{ki})] - E[dis_i|H_{ki} = 0, Pr(\mathbf{Z}_{ki})]\}. \quad (3.6)$$

O efeito médio de exposição, estimado pela abordagem PSM, é determinado pela diferença média nos resultados sobre o suporte comum, devidamente ponderados pela distribuição do *propensity score* dos participantes (CALIENDO; KOPEINIG, 2005). Para o cálculo da Equação 3.6 é preciso utilizar algum algoritmo de pareamento, pois como $Pr(\mathbf{Z}_k)$ é uma variável contínua as chances de identificar duas unidades com igual valor desse escore são ínfimas. Os métodos de pareamentos usados são os seguintes: r-vizinhos mais próximo (PSM-V), com e sem reposição, onde r refere-se a quantidade de vizinhos – nesta pesquisa os valores para r são 1 e 10; raio (PSM-R); e, Kernel (PSM-K), sendo usada a função kernel de Epanechnikov²⁶. Conforme sugerido por Rosenbaum (2010), os cálculos do efeito médio por PSM devem ser feitos por mais de uma técnica de pareamento, para analisar a manutenção dos sinais e magnitudes das estimativas. Assim, os resultados apresentados na Subseção 3.5.2 são desenvolvidos para nove diferentes formas de pareamento, onde no PSM-R e para uma parte do PSM-V são usados um nível máximo de tolerância (*caliper*) de 0,1% da distância do *propensity score*.

Duas suposições centrais são requeridas para o cálculo de τ_k . A primeira diz respeito a hipótese de existência do suporte comum ou sobreposição, no qual $0 < Pr(H_k = 1|\mathbf{Z}_k) < 1$. Outra importante suposição presume que todas as diferenças relevantes entre os indivíduos expostos e não-expostos são capturadas por \mathbf{Z}_k , onde é necessário a garantia, principalmente, da chamada hipótese de independência condicional (CIA): $Y(0), Y(1) \perp H_k | Pr(\mathbf{Z}_k)$, isto é, os fatores não-observados não afetam a decisão de exposição.

²⁶ Para maiores detalhes sobre as diferentes técnicas de pareamento, ver Rosenbaum (2010).

De acordo com [Rosenbaum \(2002\)](#), a hipótese CIA do PSM é altamente restritiva, pois admite que todas as variáveis que influenciam a decisão de exposição e os *outcomes* estão imputadas no modelo. Porém, a existência de fatores omitidos podem interferir na probabilidade de exposição da criança a um determinado fator de risco. A questão levantada por essa literatura concerne na medição da sensibilidade das estimativas em relação às características não-observadas. Trabalhos como [Gastwirth, Krieger & Rosenbaum \(1998\)](#) propõem análises de sensibilidade para examinar o tamanho do viés omitido na atribuição da variável indicadora de exposição que poderiam modificar as conclusões dos efeitos.

[Cornfield et al. \(1959\)](#), um dos pioneiros na análise de sensibilidade, conduz uma investigação sobre a relação causal entre o tabagismo e o câncer de pulmão. Através da formulação do risco relativo aparente, os autores demonstram que um fator não-observado deveria ser 9 vezes mais frequente entre os fumantes do que nos não-fumantes para o desenvolvimento do câncer, concluindo que não se pode rejeitar o fato de que o tabagismo é um hábito ofensivo a saúde e a longevidade. Nessa direção, [Gastwirth, Krieger & Rosenbaum \(1998\)](#), [Rosenbaum \(2002\)](#), [Caliendo & Kopeinig \(2005\)](#) e [Rosenbaum \(2010\)](#) sugerem a análise de sensibilidade para as estimativas do PSM para prover indícios sobre o grau de dependência dos resultados a hipótese (CIA) não-testável, tendo em vista a lógica do risco relativo aparente.

Se existe um viés omitido, dois indivíduos com idênticas características observáveis \mathbf{Z}_k têm diferentes chances de serem expostos ao fator de risco k , pois o efeito de ϵ (termo de erro estocástico) captado por γ é diferente de zero. Nessas condições, a razão de chances de dois indivíduos pareados serem expostos é expressa pela Equação 3.7

$$\frac{Pr(\mathbf{Z}_{ki})[1 - Pr(\mathbf{Z}_{kj})]}{Pr(\mathbf{Z}_{kj})[1 - Pr(\mathbf{Z}_{ki})]} = \frac{G(\mathbf{Z}_{ki}\lambda + \gamma\epsilon_i)}{G(\mathbf{Z}_{kj}\lambda + \gamma\epsilon_j)}. \quad (3.7)$$

Se os indivíduos i e j possuem iguais \mathbf{Z}_k , $G(\cdot)$ segue uma distribuição logística e $\epsilon \in (0, 1)$, [Rosenbaum \(2002\)](#) mostra que a Equação 3.7 implica nos seguintes limites para as razões de chances:

$$\Gamma^{-1} \leq \frac{Pr(\mathbf{Z}_{ki})[1 - Pr(\mathbf{Z}_{kj})]}{Pr(\mathbf{Z}_{kj})[1 - Pr(\mathbf{Z}_{ki})]} \leq \Gamma, \quad (3.8)$$

com $\Gamma = e^\gamma$.

Essa abordagem também é conhecida como limites de Rosenbaum, já que a Equação 3.8 informa o quanto as mudanças nos valores de Γ alteram a inferência acerca do efeito. Portanto, o valor de Γ mede o grau de afastamento que uma estimativa feita por PSM está livre de fatores não-observados. Neste estudo, a citada medida possibilita identificar o tamanho da influência das variáveis não-observadas sobre a probabilidade de exposição

aos fatores de risco, permitindo investigar em que medida as variáveis omitidas podem interferir na robustez dos resultados encontrados.

3.4 Base de dados

Para a execução desse estudo são utilizados os microdados da PENSE 2012, que trazem informações com abrangência nacional sobre fatores comportamentais de risco e de proteção à saúde dos estudantes. O público-alvo desse *survey* é formado por alunos do 9º ano do ensino fundamental do turno diurno de escolas públicas e particulares com no mínimo 15 alunos. Segundo o IBGE (2013), a escolha do 9º ano do ensino fundamental é realizada devido ao mínimo de escolarização necessária para responder ao questionário autoaplicável e com base na faixa etária de referência (entre 13 e 15 anos de idade) recomendada pela Organização Mundial da Saúde - OMS (*World Health Organization* - WHO) para fins de comparabilidade com pesquisas internacionais sobre o tema.

Os dados da PENSE 2012 são estruturados a partir de questionários direcionados aos alunos, que possuem informações sobre atributos do discente – com o diferencial de perguntas dirigidas às condições de saúde – e de seus familiares, e às características da escola. Essa base de dados inicialmente é composta por 109.104 estudantes do último ano do ensino fundamental pertencentes à 2.842 escolas das redes pública (municipal, estadual e federal) e privada. Após uma série de recortes amostrais e exclusão dos *missing values*, a amostra final desta pesquisa é composta por 66.844 estudantes do ensino fundamental, que representativamente correspondem a 1,9 milhão de discentes da população, em 32 estratos (27 capitais e 5 regiões compostas pelos demais municípios) e em 1.473 unidades primárias de amostragem (UPA). Essas informações do desenho amostral ajustam as estimativas calculadas ao longo da seção de resultados, com exceção para o cálculo específico da Equação 3.6, já que o grupo de controle observável é identificado por *matching* e a diferença média é ponderada pela distribuição do *propensity score*²⁷.

O desenho amostral da PENSE é feito por meio da estratificação do território nacional, no qual cada um dos 27 municípios das capitais e do Distrito Federal formam um estrato geográfico, e os demais municípios agrupados em suas regiões de origem compõem outros cinco estratos geográficos. Os municípios das capitais têm uma amostra de escolas selecionadas (unidades primárias de amostragem - UPA) e, em seguida, as turmas dessas escolas são escolhidas (unidades secundárias de amostragem - USA). Para os municípios fora da capital, as UPA são compostas por agrupamentos de municípios, as USA por unidades escolares selecionadas e as unidades terciárias de amostragem (UTA) por turmas dessas escolas (IBGE, 2013). O plano amostral da PENSE é complexo e com delineamento misto

²⁷ No cálculo do *propensity score* é feita a inclusão do desenho amostral, com a proposta de melhor inferir as variáveis que mantêm contatos com a postura de risco da criança e dispor de parâmetros para o modelo com mais proximidade das estimativas da população.

(diferentes números de estágios para seleção da amostra entre municípios de capitais e fora delas), necessitando ajustar os modelos econométricos, quando for o caso, às especificidades amostrais do banco de dados para evitar problemas de sub ou sobrestimação da variância das estimativas que afetam o processo de inferência (PESSOA; SILVA; DUARTE, 1997; SILVA; PESSOA; LILA, 2002).

Para facilitar a descrição das variáveis, as próximas três subseções exibem o detalhamento, respectivamente, da variável de resultado (atraso escolar), das variáveis referentes à exposição aos fatores de risco comportamentais à saúde e das demais covariadas.

3.4.1 Distorção idade-série

A distorção idade-série é um *outcome* educacional, relacionando-se diretamente com o abandono e a repetência escolar, bem como com o problema de evasão escolar e o ingresso tardio na escola²⁸. Para Glewwe & Jacoby (1995), o atraso escolar contradiz a teoria do capital humano, pelo fato dos custos de oportunidade do tempo da criança serem mais baixos na infância, sendo o ingresso tardio na escola mais custoso para o indivíduo e seus familiares. Outrossim, os problemas de progressão e evasão da criança na escola geram desdobramentos no valor futuro da renda esperada pelo o indivíduo, decorrentes da ampliação do *trade-off* entre as decisões de estudo e de trabalho para pessoas com atraso escolar, afetando decisivamente a formação educacional. Quando um indivíduo com 18 anos de idade ainda está cursando o último ano do ensino fundamental, a conclusão dessa etapa de ensino e a sequência para outra (ensino médio e ensino superior) possuem altos custos de oportunidade, principalmente para as pessoas mais pobres que precisam contribuir com a renda familiar (GLEWWE; JACOBY, 1995).

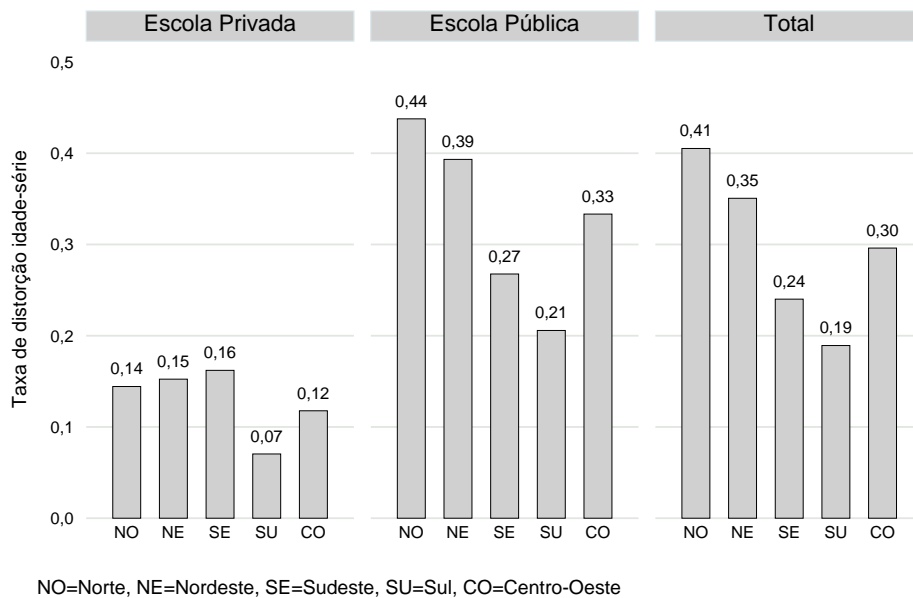
A construção da variável distorção idade-série (*dis*), doravante tratada como atraso escolar, parte do princípio que um sistema educacional seriado, como o brasileiro, há uma adequação teórica entre a série e a idade do discente. Atualmente, a idade recomendada para ingresso no ensino fundamental no Brasil é de 6 anos de idade, como prediz a Lei nº 11.274/2006, que alterou os artigos 29, 30, 32 e 87 da Lei nº 9.394/1996. Sabendo que essa etapa de ensino tem duração total de 9 anos, é possível presumir a idade adequada do estudante em cada série. Logo, o cálculo do indicador de atraso escolar (*dis*) é mostrado na Equação 3.9

$$dis_{is} = \begin{cases} 1, & \text{se } (I_{is} - I_s^e) \geq 1,5 \\ 0, & \text{c.c.} \end{cases}, \quad (3.9)$$

²⁸ Além disso, existem evidências na literatura, como em Gomes-Neto & Hanushek (1994), Gomes-Neto *et al.* (1997) e Soares (2003), que indicam uma relação negativa entre as estimativas de distorção da idade na escola e o baixo desempenho em testes padronizados

onde: $dis_{is} = 1$ indica a existência de atraso escolar, $dis_{is} = 0$ indica que o indivíduo i está frequentando a série s na idade adequada; I_{is} é a idade em anos do aluno i , obtida pelo ano e mês de nascimento, que frequenta a série s ; I_s^e é a idade recomendada para a série s , que é 14 anos para o último ano do ensino fundamental ($s = 9$). A Equação 3.9 estabelece que crianças com diferenças positivas e maiores que 1,5 ano em $(I_{is} - I_s^e)$ são enquadradas com atraso escolar, de modo que estudantes com $(I_{is} - I_s^e) = 1$ mas que nasceram no segundo semestre do ano escolar não são classificados como atrasados²⁹. Com base nesse indicador, a Figura 3.1 exibe a taxa de distorção idade-série, restringida para a amostra final deste estudo (detalhes na Seção 3.4), entre as escolas das redes privada, pública e total por agrupamento regional.

Figura 3.1 – Taxa de distorção idade-série no 9º ano do Ensino Fundamental por dependência administrativa e grandes regiões do Brasil a partir da amostra final deste trabalho



Fonte: Elaboração própria a partir dos microdados da PENSE 2012. Amostra composta de 66.844 estudantes do 9º do ensino fundamental com representatividade na população de 1,9 milhão de observações.

Verificando esses dados, é perceptível que as taxas de atraso escolar das escolas públicas, dado que as escolas privadas respondem por 20% dos alunos na amostra, definem o comportamento do nível de distorção idade-série nas regiões brasileiras. As escolas públicas, de modo geral, têm um escore três vezes pior do que o constatado nas escolas privadas, com exceção da região Sudeste que apresenta uma razão menor que dois. O destaque fica por parte das regiões Norte e Nordeste que detêm os piores indicadores na rede pública e no total. Contudo, a região Sudeste entre as escolas particulares possui

²⁹ Esse critério é baseado no dicionário de indicadores do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (2004), que considera uma margem de um ano de tolerância.

relativamente uma elevada distorção idade-série. Tais fatos sinalizam que o atraso escolar potencialmente se mostra relacionado aos fatores socioeconômicos, tendo em vista as já conhecidas disparidades socioeconômicas entre os estudantes de escolas públicas e privadas e entre as regiões Sul/Sudeste e Norte/Nordeste.

3.4.2 Indicadores de exposição aos fatores de risco comportamentais

Os tipos de fatores de risco modificáveis à saúde contemplados neste estudo são o uso do cigarro, o consumo de bebidas alcoólicas e o sedentarismo/má alimentação (medido pelo sobrepeso). Em relação ao último indicador, destaca-se que a PENSE apenas possui informações sobre as condições correntes do consumo de alimentos e de práticas de atividades físicas. Por isso, que as informações sobre o índice de massa corporal³⁰ (IMC) são usadas como *proxies* para o estoque de hábitos não saudáveis referentes à alimentação e práticas de atividades físicas. Crianças com valor de IMC acima do 85º quantil da distribuição do IMC específica por idade-gênero são denotadas como estudantes com excesso de peso.

As condições nutricionais, ditadas pelo IMC, captam o estoque passado dos investimentos em saúde, dentre eles a adoção de hábitos de consumo não saudáveis (dieta inadequada) e falta de práticas de atividades físicas³¹. Essa abordagem que obedece a um padrão empírico e específico para cada idade e gênero, dentro de intervalos quantílicos definidos, é recomendada pela OMS para crianças e adolescentes (KUCZMARSKI *et al.*, 2002). Os limites (valores mínimos) do IMC que definem os estudantes com sobrepeso para a amostra considerada e o critério de classificação mencionado estão expressos a seguir³²:

- Meninas: com 13 anos ou menos, 24,89 kg/m^2 ; com 14 anos, 25,31 kg/m^2 ; com 15 anos, 25,92 kg/m^2 ; com 16 anos, 26,95 kg/m^2 ; com 18 anos, 27,74 kg/m^2 ; com 18 anos ou mais, 29,30 kg/m^2 ;
- Meninos: com 13 anos ou menos, 24,77 kg/m^2 ; com 14 anos, 24,80 kg/m^2 ; com 15 anos, 25,39 kg/m^2 ; com 16 anos, 26,37 kg/m^2 ; com 18 anos, 27,46 kg/m^2 ; com 18 anos ou mais, 27,73 kg/m^2 .

³⁰ O cálculo desse indicador é dado por: $IMC = peso/(altura^2)$, com o peso expresso em quilograma (kg) e a altura em metro (m).

³¹ Abrevaya & Tang (2010) realçam que as condições nutricionais das crianças são determinadas primariamente por fatores genéticos e pelo ambiente comum (hábitos saudáveis, questões locacionais, *peer-effects*, entre outros). No entanto, para eles a taxa de crescimento de indivíduos com excesso de peso nos últimos anos indica que o fator ambiente comum vem assumindo um maior grau de importância.

³² Como à definição do excesso de peso utiliza a distribuição por quantis do IMC, as crianças com idade inferior a 13 anos e superior a 18 anos tiveram que ser avaliadas em conjunto, respectivamente, com alunos de 13 anos de idade e de 18 anos de idade, devido à insuficiência amostral.

Quanto às medidas de exposição ao consumo de álcool e do cigarro, o questionário do estudante aplicado na PENSE dispunha de informações sobre a idade de primeiro contato com tais produtos e a indicação do consumo deles nos últimos 30 dias. Os alunos que declararam o primeiro contato com bebida alcoólica ou com cigarro com mais de 14 anos de idade foram desconsiderados da amostra, pois a primeira exposição teria ocorrido em uma faixa etária acima da idade teórica para o 9º ano do ensino fundamental. Este trabalho usa duas medidas indicativas sobre esses comportamentos, uma relacionada ao tempo de experiência da utilização do produto e outra que combina essa experiência com o uso recente³³. Assim, esses indicadores estão divididos em: definição 1 – experiência com o produto com idade igual ou anterior aos 14 anos, doravante denotada apenas por *experiência*; definição 2 – experiência combinada com o uso nos últimos 30 dias do período da aplicação do questionário da PENSE, denominada a partir de agora de *experiência e uso*.

A Figura 3.2 exibe a proporção de alunos expostos aos fatores de risco à saúde em todos os casos avaliados, para a amostra final deste estudo. Os dados revelam um número preocupante de crianças expostas ao uso de bebidas alcoólicas com idade inferior ou igual a 14 anos de idade (66% na definição 1 e 24% na definição 2). Quanto ao cigarro, 18% dos alunos declaram o consumo do produto com 14 anos ou menos de idade e 4% destes consumiram o produto no último mês. Por sua vez, o excesso de peso contempla aproximadamente 16% dos indivíduos avaliados³⁴.

3.4.3 Demais variáveis

Considerando a função de produção educacional, o vetor \mathbf{X} é formado por: atributos da criança (gênero, raça, trabalho infantil) e de seus familiares (núcleo familiar, nível socioeconômico, quantidade de membros na família); características da escola (total de salas de aula, dependência administrativa pública, existência de conselho escolar³⁵, indicador de infraestrutura e média do nível socioeconômico dos alunos); fatores locacionais (capital e grande região geográfica).

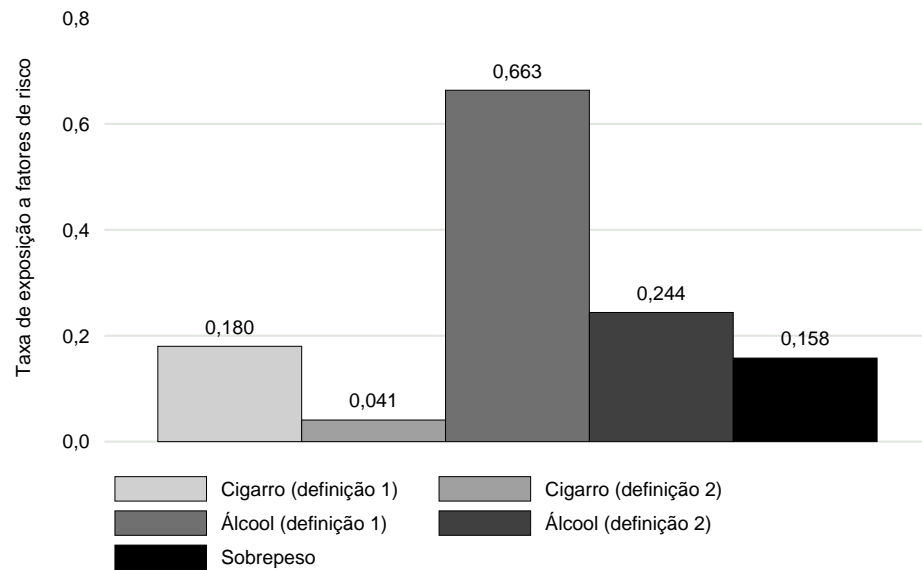
Consonante à estratégia empírica adotada, além do vetor \mathbf{X} , o vetor de características observáveis \mathbf{Z}_k é usado na estimação do *propensity score* para cada fator de exposição k . O vetor \mathbf{Z}_k é composto pelo vetor \mathbf{X} e por variáveis situacionais relativas ao ambiente familiar e escolar, a saber: na família – indicativos de pais fumantes, de reação da família quanto ao uso de bebida alcoólica pela criança e de promoção de hábitos alimentares

³³ Entende-se como uso recente, o consumo do produto nos últimos 30 dias de referência da PENSE.

³⁴ O cálculo dos limites do IMC para a definição do excesso de peso é desenvolvido em relação aos 98.080 estudantes que possuíam dados sobre altura e peso, por isso que a frequência relativa de indivíduos acima do quantil 85 é superior a 15%, uma vez que a amostra final possui uma quantidade menor de observações.

³⁵ O conselho escolar presume a participação dos professores, funcionários, pais e alunos no processo decisório da gestão da escola em conjunto com a direção escolar.

Figura 3.2 – Taxa de exposição aos fatores de risco à saúde por parte de estudantes do 9º ano do Ensino Fundamental no Brasil a partir da amostra final deste trabalho



Fonte: Elaboração própria a partir dos microdados da PENSE 2012. Amostra composta de 66.844 estudantes do 9º do ensino fundamental com representatividade na população de 1,9 milhão de observações.

saudáveis (consumo de frutas e legumes); na escola – taxa de prevalência de exposição ao fator de risco k entre os alunos. As estatísticas descritivas dessas variáveis estão detalhadas na Tabela B.3 no Apêndice.

O nível socioeconômico (NSE) e o indicador de infraestrutura escolar (IEE) são estimados através da Teoria da Resposta ao Item (TRI), abordagem estatística que ajusta um conjunto de respostas conforme as propriedades de cada item avaliado para estimar traços latentes, como habilidades cognitivas, qualidade de infraestrutura, nível socioeconômico etc. (LINDEN; HAMBLETON, 2010). O traço latente do NSE e do IEE captura, respectivamente, o nível socioeconômico do estudante e a qualidade dos recursos estruturais (serviços e infraestrutura) da escola. Para a estimação do traço latente é usado um algoritmo Bayesiano, conhecido como estimador *expected a posteriori* (EAP).

A seguir são apresentadas breves explicações sobre a construção dessas duas variáveis:

- NSE representa um indicador que mede o patrimônio e outras condições socioeconômicas dos alunos. Esse indicador está baseado em Soares & Andrade (2006) e Alves & Soares (2012) que usam o modelo da Teoria da Resposta ao Item (TRI), visto que bases de dados (como a da PENSE) não dispõem de informações sobre o rendimento familiar, bem como o NSE consegue sintetizar em uma única medida uma série de itens associados as condições socioeconômicas. Todos os itens que medem informações

diretas e indiretas sobre a escolaridade dos pais e a renda da família são considerados na estimação do NSE. No total, dez itens são incorporados na mensuração do NSE do aluno (descritos com mais detalhes na Tabela B.1 no Apêndice), com destaque para as variáveis de escolaridade dos pais, de patrimônio e de serviços disponíveis na casa da criança. Segundo as informações presentes nessa tabela, a grande parte dos estudantes (65%) só dispõe de um banheiro em casa, 42% das mães e 47% dos pais nunca estudaram ou não concluíram o ensino fundamental e apenas 10% das casas possuem empregado doméstico. Os parâmetros estimados pela TRI para cada um dos itens pontuam que aqueles relacionados a disponibilidade de computador e acesso à internet em casa são os itens com maior peso de discriminação entre indivíduos com alto e baixo NSE, enquanto a posse de moto é o item que menos distingue. O modelo logístico policotômico de dois parâmetros da TRI (unidimensional e com um sistema de crédito parcial generalizado) é o escolhido para estimação do NSE, devido ao fato dos itens referentes à escolaridade e a quantidade de banheiros terem mais de duas alternativas ordenadas possíveis. O NSE está normalizado³⁶ para o intervalo entre 0 e 1. Todavia, o valor estimado desse traço latente originalmente se situa entre $-\infty$ e $+\infty$.

- IEE considera onze perguntas existentes no questionário da escola sobre a disponibilidade de recursos (equipamentos, ambientes e serviços). Esses itens são: biblioteca, laboratório de informática, computadores em sala de aula, internet na escola, cantina, quadra de esportes, quadra de esporte com cobertura, pista de corrida, piscina e vestiário (ver Tabela B.2 no Apêndice). Os recursos com maior escassez entre as escolas da amostra são pista de corrida (2,6%), piscina (14%) e acesso a computadores em sala de aula (28%), enquanto os mais recorrentes são biblioteca (91%), laboratório de informática (87%) e acesso à internet (86%). Quanto aos itens que mais discriminam as escolas em termos de IEE, em consonância com os parâmetros estimados, destacam-se a existência de quadra de esportes e piscina. Por outro lado, itens que tratam da existência de computadores na escola e cantina são os que menos discriminam. O traço latente para o IEE é calculado usando a versão dicotômica unidimensional da TRI com dois parâmetros, tendo por base Soares-Neto *et al.* (2013), no qual o indicador normalizado (transformação idêntica a promovida no NSE) varia de 0 a 1.

3.5 Resultados

Os resultados encontram-se divididos em três partes: estimação da função de distorção idade-série para exploração inicial dos resultados e primeiras indicações da

³⁶ A normalização é realizada pela seguinte expressão: $\theta_{iN} = \frac{\theta_i - \theta_{min}}{\theta_{max} - \theta_{min}}$, onde θ_{iN} representa o traço latente normalizado do indivíduo $i \in (0, 1)$ e $\theta_i \in (-\infty, +\infty)$ é o traço latente do indivíduo i .

associação dos hábitos não saudáveis com o *outcome* escolar; cálculo do efeito médio da exposição por meio de técnicas de pareamento; análise da sensibilidade das estimativas do PSM aos fatores não-observados.

3.5.1 Exploração inicial: Determinantes da atraso escolar

As estimativas dos determinantes do atraso escolar, controlando para os atributos dos alunos (incluindo os fatores de risco comportamentais) e de suas famílias, as características da escola e os fatores locacionais, estão exibidos na Tabela 3.1. Constan nessa tabela nove especificações de modelo, que se diferenciam pela inclusão e interação entre a exposição do aluno aos fatores de risco à saúde modificáveis. Inicialmente, os sinais, a magnitude e a significância estatística das variáveis explicativas mostram o mesmo padrão para todas as especificações, onde de forma geral os modelos conseguem classificar corretamente 73% dos casos, isso assumindo um *cut-off* padrão de 50% na probabilidade de atraso escolar.

Quanto ao primeiro grupo de variáveis, atributos gerais do aluno e da família, pontua-se que as variáveis referentes ao gênero, trabalho e a quantidade de membros da família elevam as chances de atraso escolar, ao passo que alunos de cor branca, com maior nível socioeconômico (NSE) e que co-residem com ambos os pais tendem a ter um melhor desempenho. Esses indicativos corroboram a literatura empírica internacional e nacional, que enfatizam o papel do perfil socioeconômico do discente no rendimento escolar. Observando os efeitos marginais³⁷ na Tabela B.4 no Apêndice estimados para esse vetor de variáveis, o NSE do aluno tem uma importante contribuição na probabilidade de atraso escolar. Um incremento unitário no indicador socioeconômico resulta em uma redução em torno de 26% na probabilidade de estar atrasado na escola.

As unidades escolares com melhores atributos em termos de gestão escolar democrática (reflexo da existência de conselho escolar), de qualidade de infraestrutura e de melhor *background* socioeconômico de seus alunos reduzem as chances do discente apresentar distorção idade-série. Já o tamanho da escola, ditado pelo total de matrículas, não é estatisticamente significativo. O efeito marginal³⁸ (*Emg*) do IEE é de aproximadamente -13% na probabilidade de atraso escolar e o NSE do conjunto dos alunos da escola, os chamados *peer-effects*, possui uma relação importante com o desempenho do discente ($Emg \approx -38\%$).

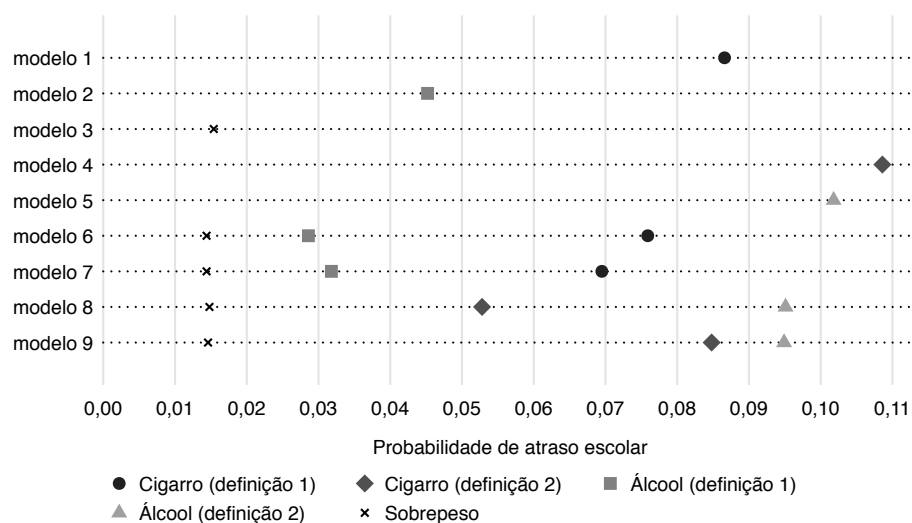
³⁷ O efeito marginal para variáveis discretas foi calculado pela Equação 3.4 e para as variáveis contínuas pelo o efeito parcial médio (EPM).

³⁸ A Tabela B.4 no Apêndice exibe em detalhes as estimativas de efeito marginal de todas as covariadas.

O controle para características do aluno e da escola retiram os efeitos da rede de dependência administrativa, porque as escolas públicas brasileiras possuem alunos com menos *background* socioeconômico, $-40\% < Emg < -38\%$, e pior infraestrutura, $Emg \approx -33\%$, do que as escolas da rede particular de ensino. No que tange aos fatores atrelados às características locais da unidade escolar na Tabela 3.1, nenhum dos coeficientes se mostram estatisticamente significativos, dada principalmente as altas variâncias dessas estimativas quando se pondera o desenho amostral do *survey* em análise.

Dedicando agora para uma primeira análise da relação entre os fatores de risco comportamentais à saúde e o desempenho na escola, identifica-se uma relação direta entre o baixo indicador educacional do aluno e à exposição aos fatores de risco. Pelos dados da Tabela 3.1, crianças expostas às bebidas alcoólicas, ao cigarro e ao excesso de peso possuem mais chances de estarem atrasadas na escola³⁹. A princípio essa relação corrobora a hipótese dos achados empíricos sobre as condições de saúde física e psicológica vistos na literatura internacional. Para melhor dimensionar os efeitos médios desses hábitos prejudiciais à saúde sobre a *performance* na escola, a Figura 3.3 ilustra tais efeitos para cada uma das nove diferentes especificações.

Figura 3.3 – Efeito médio dos fatores de risco comportamentais à saúde sobre o atraso escolar – não condicionado ao contrafactual observável



Fonte: Elaboração própria a partir dos microdados da PENSE 2012. Todos os coeficientes são significativos estatisticamente a pelo menos 10%.

Os alunos expostos ao excesso de peso, ao cigarro e as bebidas alcoólicas, de acordo com o modelo (9) na Figura 3.3, apresentam respectivamente uma probabilidade, em

³⁹ No caso do modelo (9), as crianças que adotam isoladamente apenas uma das posturas de risco (beber ou fumar) apresentam aproximadamente um coeficiente de 0,52 cada, contudo os alunos expostos simultaneamente aos dois produtos possuem um coeficiente de 0,74. Assim, adotar as duas posturas não saudáveis em conjunto possui um efeito ainda maior nas chances de atraso escolar, sendo que o sinal negativo da interação entre essas exposições sugere uma interdependência dos efeitos.

Tabela 3.1 – Resultados da estimação do modelo logit para a probabilidade de atraso escolar dos estudantes no Brasil 2012

Covariadas	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Atributos do aluno e da família									
Gênero (homem=1)	0,5195***	0,5331***	0,5188***	0,5221***	0,5381***	0,5291***	0,5294***	0,5392***	0,5394***
Raça (Branca=1)	-0,1548***	-0,1605***	-0,1597***	-0,1566***	-0,1568***	-0,1557***	-0,1558***	-0,1553***	-0,1554***
Trabalha	0,5054***	0,5236***	0,5502***	0,5278***	0,4826***	0,4930***	0,4926***	0,4745***	0,4747***
Núcleo familiar biparental	-0,4014***	-0,4142***	-0,4351***	-0,4223***	-0,4011***	-0,3920***	-0,3922***	-0,3963***	-0,3965***
NSE do aluno	-1,4456***	-1,4988***	-1,4506***	-1,4522***	-1,5065***	-1,4744***	-1,4745***	-1,5018***	-1,5028***
Tamanho da família	0,0801***	0,0838***	0,0820***	0,0804***	0,0809***	0,0813***	0,0814***	0,0800***	0,0801***
Atributos da escola									
Total de matrículas (em mil)	0,0488	0,0509	0,0539	0,0511	0,0535	0,0467	0,0468	0,0513	0,0509
Rede pública	-0,0921	-0,0698	-0,0634	-0,0741	-0,0883	-0,0923	-0,0927	-0,0916	-0,0919
Conselho Escolar	-0,2578***	-0,2522***	-0,2513***	-0,2536***	-0,2531***	-0,2575***	-0,2574***	-0,2539***	-0,2547***
Índice de infraestrutura (IEE)	-0,7190***	-0,7121***	-0,7072***	-0,7122***	-0,7224***	-0,7183***	-0,7181***	-0,7213***	-0,7219***
NSE da escola	-2,1671***	-2,1781***	-2,1082***	-2,1257***	-2,1595***	-2,2067***	-2,2081***	-2,1679***	-2,1664***
Fatores regionais									
Capital	-0,3283	-0,3398	-0,3433	-0,3374	-0,3422	-0,3269	-0,327	-0,3384	-0,3379
Região Norte	0,1508	0,15	0,1384	0,1491	0,167	0,1602	0,1603	0,174	0,174
Região Nordeste	-0,114	-0,1219	-0,1237	-0,1141	-0,1201	-0,1086	-0,1087	-0,1102	-0,11
Região Sul	-0,4072	-0,3748	-0,3469	-0,3673	-0,3959	-0,4164	-0,416	-0,4021	-0,4027
Região Centro-Oeste	0,0321	0,0503	0,0587	0,0509	0,0468	0,0341	0,0346	0,0477	0,0478
Exposição aos fatores de risco									
Experiência: cigarro (def.1)	0,4585***					0,4046***	0,2921***		
Experiência: álcool (def.1)		0,2553***				0,1611*	0,1552*		
Sobrepeso			0,0848*			0,0805*	0,0804*	0,0826*	0,0818*
Experiência e uso: cigarro (def.2)								0,2834***	0,5218***
Experiência e uso: álcool (def.2)								0,5081***	0,5201***
Experiência: cigarro e álcool					0,5423***				
Experiência e uso: cigarro e álcool							0,1208		-0,3017***
Constante	1,3905*	1,3098	1,4317*	1,4337*	1,3632*	1,2948	1,2988	1,3451*	1,3423*
Observações	66.844	66.844	66.844	66.844	66.844	66.844	66.844	66.844	66.844
População	1.897.670	1.897.670	1.897.670	1.897.670	1.897.670	1.897.670	1.897.670	1.897.670	1.897.670
Estratos	32	32	32	32	32	32	32	32	32
UPA	1.473	1.473	1.473	1.473	1.473	1.473	1.473	1.473	1.473
Count-R2 (%)	73,30	73,21	73,09	73,24	73,39	73,28	73,30	73,47	73,45

Fonte: Elaboração própria a partir dos microdados da PENSE 2012.

Nota: As estimativas são desenvolvidas considerando, além do peso amostral, a estratificação e a UPA do plano amostral.

Legenda: *p-valor<0,10; **p-valor<0,05; ***p-valor<0,01.

média, de atraso escolar superior em 1,5%, 8,5% e 9,5% no comparativo com a situação de não-exposição. Ao longo de cada uma das especificações disponíveis, os efeitos apresentam pequenas alterações de magnitudes (em especial para definição 1 da exposição ao álcool e definição 2 da exposição ao cigarro), mas os sinais dos coeficientes permanecem mantidos. Os efeitos médios variam: de 7,0% a 8,7% na definição 1 e de 5,3% a 10,9% na definição 2, para o cigarro; de 2,9% a 4,5% na primeira definição e de 9,5% a 10,2% na segunda, para o consumo de bebida alcoólica; de 1,4% a 1,5%, para o excesso de peso. No tocante ao NSE dos alunos, uma pergunta de interesse para as políticas públicas que pode ser colocada neste ponto é: como as diferenças nos resultados escolares de alunos expostos e não-expostos aos fatores de risco à saúde se comportam em diferentes níveis socioeconômicos? Assim, a Figura 3.4 expressa as diferenças na probabilidade de atraso escolar provocado pela exposição aos fatores de risco, mediante diferentes níveis socioeconômicos. As estimativas são elaboradas com base nos resultados da última especificação do modelo logit dispostos na Tabela 3.1, com o NSE variando do seu valor mínimo (0) ao seu valor máximo (1) por meio de incrementos de 0,10 unidades na citada variável.

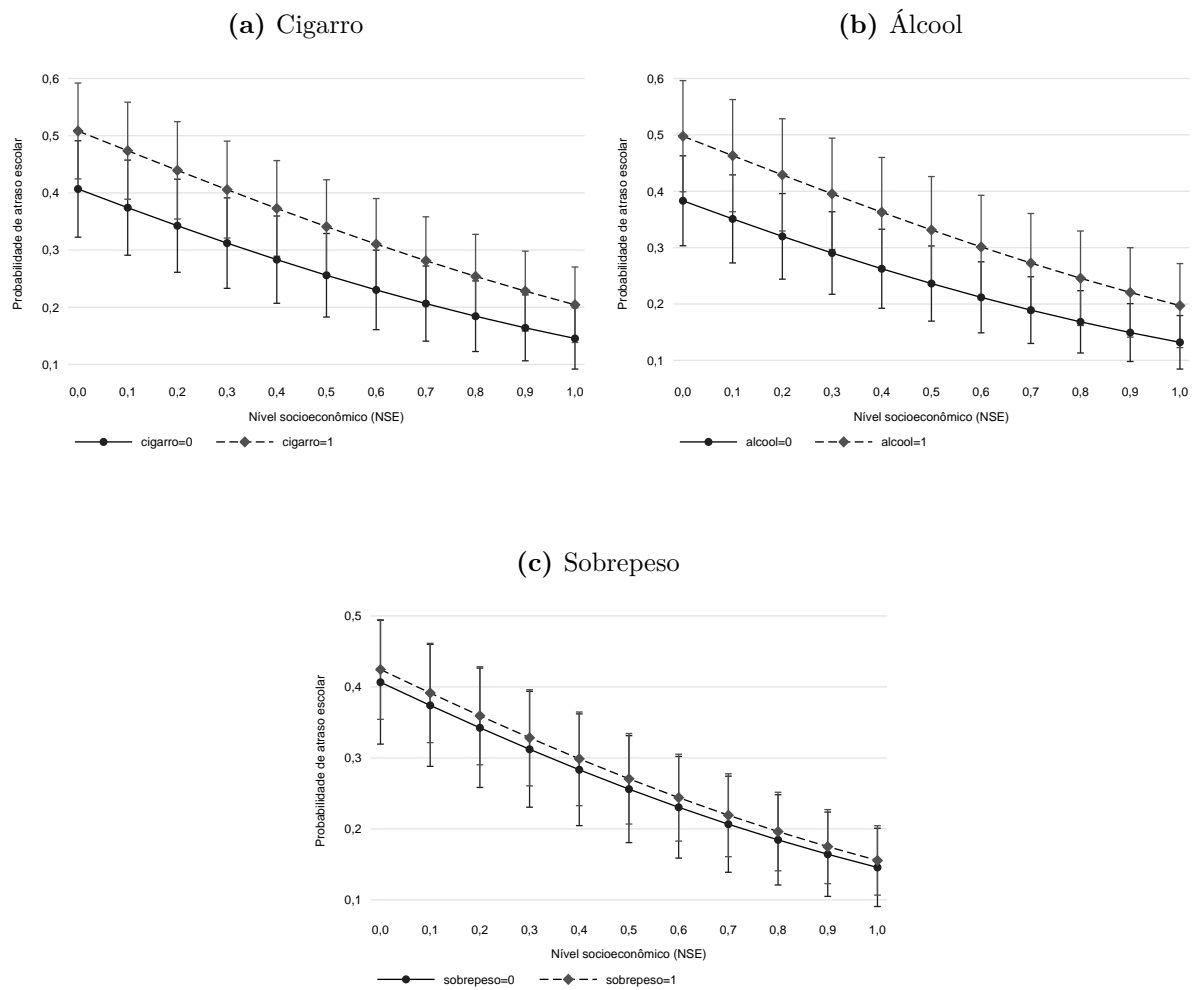
Por meio da Figura 3.4, as diferenças nas probabilidades de atraso escolar entre os expostos e não-expostos são maiores para os alunos mais pobres. Essa constatação pode ser inicialmente visualizada na Figura 3.4a, onde estudantes usuários do cigarro com menor nível de NSE têm uma diferença na probabilidade de distorção idade-série de 10 pontos percentuais (p.p.), enquanto que para os indivíduos com maior NSE essa diferença cai para 5,9 p.p. Em igual direção ocorre com as bebidas alcoólicas, onde essa diferença de resultado em decorrência desse tipo de consumo diminui gradativamente na medida em que o *background* socioeconômico melhora – o efeito para o nível mínimo de NSE é de 11,5 p.p. e para o maior nível de 6,5 p.p. Em relação ao excesso de peso na Figura 3.4c, as diferenças entre as probabilidades de atraso também são superiores nos níveis iniciais de NSE, em que, por exemplo, o efeito médio da exposição é de 1,8 p.p., no primeiro nível, e de 1 p.p. no último.

3.5.2 Efeito médio da exposição usando PSM

Esta subseção explicita os resultados obtidos pela abordagem PSM, para verificar se os efeitos médios da exposição são mantidos. Os primeiros achados na subseção anterior sinalizam que os alunos expostos aos fatores de risco à saúde tendem a apresentar uma pior *performance* educacional. Especificamente a essas estimativas, a preocupação inicial em termos da inferência dos resultados diz respeito a garantia da ortogonalidade entre a variável indicadora da exposição e os fatores omitidos. Nesse cenário, é recomendável a consideração de outra abordagem estatística que possibilite outra forma de comparabilidade entre as possíveis diferenças nos resultados educacionais provocadas pelos fatores de risco.

Pensando nas características observáveis que podem afetar a decisão de exposição,

Figura 3.4 – Efeito dos fatores de risco comportamentais à saúde sobre o atraso escolar por NSE no Brasil 2012 – não condicionado ao contrafactual observável



Fonte: Elaboração própria a partir dos microdados da PENSE 2012. Todos os coeficientes são significativos estatisticamente a pelo menos 1%.

Nota: Estimativas desenvolvidas através dos parâmetros do modelo (9) na Tabela 3.1.

esse instrumental estatístico visa identificar um grupo de estudantes não-expostos ao fator de risco k que guardem idênticos atributos \mathbf{Z}_k , de modo que a diferença nos resultados escolares possam ser atribuídas à exposição ao fator de risco. Para a construção do efeito médio sobre o grupo exposto ao fator k , denotado como τ_k , é preciso inicialmente estimar o *propensity score* $Pr(\mathbf{Z}_k)$ e, em seguida, utilizar algoritmos de pareamento para o cálculo do τ_k . A Tabela 3.2 exhibe os coeficientes usados na estimação das probabilidades de exposição ao fator de risco à saúde⁴⁰ k (com os números entre parênteses indicando a definição usada

⁴⁰ A escolha por cada hábito comportamental de risco é modelada de forma independente, pois a relação entre as diferentes exposições não são altas ou estatisticamente diferente de zero, tal como encontrada em outros estudos como Cutler & Glaeser (2005). A Figura B.1 no Apêndice ilustra que os termos residuais das equações para os determinantes de cada hábito não possuem relações bem definidas, quando se considera o ajustamento das curvas por uma regressão não-paramétrica a partir da abordagem *Linear Locally Estimation*.

para o consumo de álcool e cigarro, em conformidade com a descrição na parte sobre a base de dados).

Tabela 3.2 – Resultados da estimação dos coeficientes para o *propensity score* por tipo de exposição ao fator de risco à saúde da criança no Brasil 2012

Covariadas	Cigarro(1)	Cigarro(2)	Álcool(1)	Álcool(2)	Sobrepeso
Gênero (homem=1)	0,0199	-0,1318	-0,2755***	-0,1899***	-0,0161
Raça (Branca=1)	-0,0813**	-0,1998***	0,0268	-0,026	-0,0017
Trabalha	0,6392***	0,8592***	0,5302***	0,6431***	0,0593
Núcleo familiar biparental	-0,4570***	-0,5101***	-0,3853***	-0,3221***	-0,012
NSE do aluno	-0,1469	-0,0177	0,7808***	0,3957**	-0,107
Tamanho da família	0,0233**	0,0745***	-0,0233***	0,0226**	0,0054
Pai fumante	0,3716***	0,3458***			
Mãe fumante	0,3766***	0,3009***			
Reação da família ao álcool			-1,0998***	-1,3870***	
Frequência de consumo: legumes					0,0056
Frequência de consumo: frutas					-0,0119**
Tamanho da escola	0,0237	0,1037***	0,0696**	0,0501***	0,0502***
Rede pública	0,0097	0,0901	-0,045	-0,0564	-0,0093
Conselho Escolar	-0,0716***	-0,0684	0,0042	0,0298	0,0142
Índice de infraestrutura (IEE)	-0,0244	-0,3697***	-0,1477	-0,026	-0,0208
NSE da escola	0,6864***	1,2920***	1,1731***	0,0742	0,1362
Prevalência de exposição na escola ^a	5,9415***	14,3350***	2,8594***	5,2736***	6,9877***
Capital	0,0093	-0,0277	-0,0433	0,0474*	0,0131
Região Norte	-0,0842*	-0,2639**	-0,1027	-0,0482	-0,0506**
Região Nordeste	-0,0671	-0,3272***	-0,0269	0,0675	-0,0469***
Região Sul	0,0536***	0,1715***	0,3415***	0,0571	0,0045
Região Centro-Oeste	0,001	-0,0325	0,0395	0,0051	-0,0248
Constante	-2,9883***	-4,7274***	0,6090***	-1,5817***	-2,9053***
Observações	66.844	66.844	66.844	66.844	66.844
População	1.897.670	1.897.670	1.897.670	1.897.670	1.897.670
Estratos	32	32	32	32	32
UPA	1.473	1.473	1.473	1.473	1.473
Count-R2 (%)	82,16	95,85	67,85	77,28	84,19
Taxa de suporte (%)	47,94	64,24	30,49	37,47	57,75
Previsão correta de $H_k = 0$ (%)	97,89	99,83	17,75	96,26	99,86
Previsão correta de $H_k = 1$ (%)	9,69	2,05	92,75	19,83	0,60

Fonte: Elaboração própria a partir dos microdados da PENSE 2012.

Nota: As estimativas são desenvolvidas considerando, além do peso amostral, a estratificação e a UPA do plano amostral.

Legenda: *p-valor<0,10; **p-valor<0,05; ***p-valor<0,01.

^a Em consonância com o fator de risco considerado. Por exemplo: no modelo referente ao excesso de peso, essa variável reflete a taxa de alunos na escola enquadrados nessa definição.

Analisando os indicadores de grau de ajustamento dos modelos estimados na Tabela 3.2, a proporção de casos corretamente previstos é de no mínimo 67,85% para o modelo que traz como variável dependente a decisão de exposição ao consumo de álcool (definição 1), apontando que em todos os modelos as classificações corretas são bem superiores as ocorrências classificadas erroneamente. Pela taxa de previsão correta para indivíduos não-expostos ($H_k = 0$) e expostos ($H_k = 1$), os modelos conseguem fazer previsões de estudantes para cada um dos cenários de exposição, independentemente da distribuição da frequência amostral de cada uma das variáveis dependentes, que em geral são muito concentradas em $H_k = 0$. Entretanto, a pequena taxa de classificações corretas de $H_k = 1$ no modelo para excesso de peso alerta para possíveis problemas de especificações (visto

que o questionário da PENSE não dispõe, por exemplo, de questões acerca das condições nutricionais dos pais, sendo fatores omitidos da modelagem). Quanto ao modelo cigarro (2), essa taxa pode ser reflexo da altíssima concentração de indivíduos não-expostos na amostra (96%).

Através da taxa de suporte, que expõe a proporção de indivíduos classificados como expostos pelo modelo mas que na verdade não adotaram posturas de risco, nota-se que, para o *cut-off* padrão de 50%, todos os modelos possuem ocorrências de indivíduos candidatos ao pareamento. Essa interseção entre estudantes não-expostos que poderiam ser classificados como expostos e indivíduos corretamente demarcados como $H_k = 1$ sugere que uma das condições necessárias para o PSM está sendo respeitada, que é a hipótese de suporte comum. Outro indicativo da existência de um área de suporte pode ser visualizada na Figura B.2 no Apêndice, elaborada para cada um dos fatores de risco k em suas diferentes definições. Na citada ilustração, as curvas de densidade do *propensity score* entre os dois grupos para cada um desses fatores têm uma região de sobreposição, ratificando a possibilidade de compatibilização entre as unidades para o processo de pareamento.

As variáveis na Tabela 3.2 relativas ao núcleo familiar do aluno, trabalho, tamanho da escola, taxa de prevalência do fator de risco na escola e fatores situacionais específicos da família perante ao sobrepeso (captado pela prática alimentar) e uso de bebida alcoólica e cigarro apresentam praticamente a mesma direção de efeitos sobre a probabilidade de exposição k . No vetor do contexto escolar, as escolas de maior porte em termos de matrículas geram uma repercussão positiva na decisão do aluno assumir posturas de risco à saúde, em que tal fato pode estar refletindo uma maior influência dos colegas de escola, devido as chances de maior heterogeneidade de comportamentos não-saudáveis entre os alunos. Em igual sentido mas com magnitude maior, as escolas detentoras de maiores taxas de prevalência de alunos expostos dispõem de um cenário de maior incentivo para um determinado estudante ser influenciado pelos seus colegas. Não é por acaso que esse *peer-effect* exibe um elevado efeito marginal na probabilidade do discente ser exposto a uma escolha prejudicial a sua saúde. Em suma, os sinais dos efeitos das variáveis contextuais das escolas estão condizentes com outros trabalhos, a exemplo de Veronneau, Trempe & Paiva (2014), que apresentam evidências sobre o papel de relevo dos colegas na formação psicossocial da criança e dos efeitos desfavoráveis de amizades com pares desviantes.

As variáveis situacionais da família também são importantes para a tomada de decisão da criança, visto que filhos de pais fumantes são mais propensos à utilização do cigarro. Quanto ao consumo de bebidas alcoólicas, as famílias que reagem de forma muito contrária a adoção desse comportamento pela criança reduzem em 19%, para o modelo álcool (1), e 28%, para álcool (2), a probabilidade do aluno ter esse tipo de postura. Por sua vez, as boas práticas alimentares promovidas pela família diminuem as chances do discente obter essa classificação, pois as crianças que consomem com mais frequência frutas

dispõem de menos chances de terem excesso de peso. Logo, seja no ambiente familiar ou na escola, a decisão da criança se expor a um fator de risco, sobretudo, ao álcool e ao cigarro, é diretamente proporcional aos fatores contextuais vivenciados por ela, o que por si só tais aspectos geram implicações de políticas.

Após a estimação do *propensity score* e a identificação do grupo de controle com similares características observadas, a Tabela 3.3 contém os resultados do efeito médio sobre o grupo exposto baseado em nove formas de pareamento, ditadas pelos PSM-V e suas variantes, PSM-R e PSM-K, em relação ao atraso escolar dos discentes.

Tabela 3.3 – Efeito dos fatores de risco comportamentais à saúde sobre o atraso escolar por diferentes técnicas de pareamento no Brasil 2012 – condicionado ao contrafactual observável

Modelo	Cigarro(1)	Cigarro(2)	Álcool(1)	Álcool(2)	Sobrepeso
PSM-V (CR)					
Efeito médio	0,0855***	0,1258***	0,0411***	0,0900***	0,0237***
Erro-padrão	0,00800	0,01303	0,00494	0,00641	0,00695
PSM-V (CR, SC)					
Efeito médio	0,0854***	0,1259***	0,0417***	0,0901***	0,0237***
Erro-padrão	0,00665	0,01372	0,00523	0,00654	0,00696
PSM-V (CR, <i>Caliper</i>)					
Efeito médio	0,0850***	0,1255***	0,0418***	0,0908***	0,0239***
Erro-padrão	0,00585	0,01430	0,00506	0,00681	0,00627
PSM-V (CR, <i>Caliper</i>, SC)					
Efeito médio	0,0850***	0,1255***	0,0419***	0,0908***	0,0239***
Erro-padrão	0,00691	0,01319	0,00469	0,00653	0,00782
PSM-V (SR)					
Efeito médio	0,0844***	0,1240***	0,0404***	0,0935***	0,0239***
Erro-padrão	0,00646	0,01224	0,00386	0,00512	0,00654
PSM-V (SR, <i>Caliper</i>)					
Efeito médio	0,0850***	0,1244***	0,0358***	0,0872***	0,0244***
Erro-padrão	0,00563	0,01187	0,00406	0,00544	0,00625
PSM-V10 (SR, <i>Caliper</i>)					
Efeito médio	0,0836***	0,1272***	0,0395***	0,0851***	0,0168***
Erro-padrão	0,00519	0,01059	0,00488	0,00577	0,00534
PSM-R					
Efeito médio	0,0847***	0,1283***	0,0413***	0,0865***	0,0158***
Erro-padrão	0,00504	0,00986	0,00443	0,00486	0,00512
PSM-K					
Efeito médio	0,0869***	0,1366***	0,0423***	0,0872***	0,0159***
Erro-padrão	0,00496	0,00963	0,00421	0,00477	0,00505

Fonte: Elaboração própria a partir dos microdados da PENSE 2012.

Nota: Erro-padrão calculado através de *bootstrap* com 100 reposições. No pareamento realizado por PSM-K o valor da janela (*bandwidth*) usado foi de 0,06.

Legenda: CR=com reposição; SR=sem reposição; SC=suporte comum; *Caliper*, nível de tolerância, de 0,1%. O SC impõe um suporte comum que desconsidera observações de tratamento com *propensity score* maior que o valor máximo do escore dos controles ou menor do que o valor mínimo.

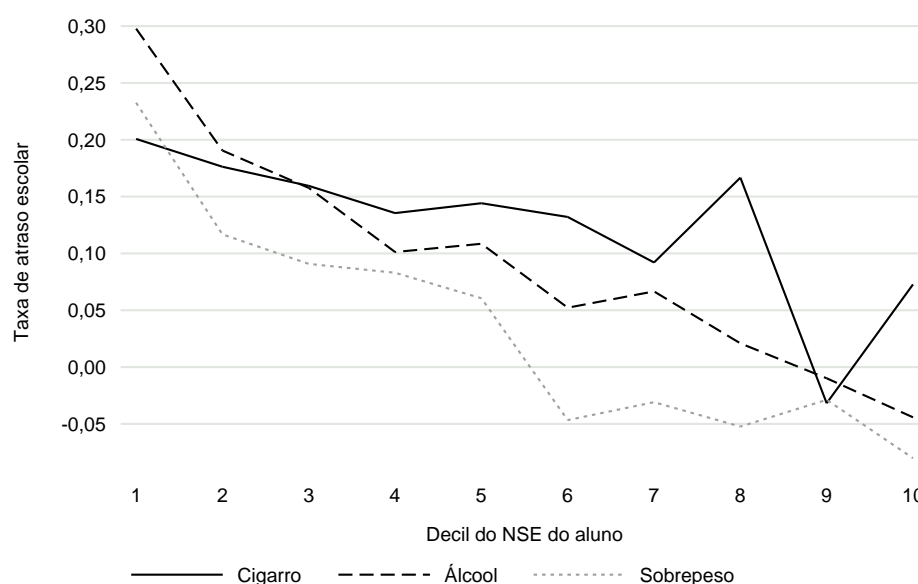
*p-valor<0,10; **p-valor<0,05; ***p-valor<0,01.

Os resultados na Tabela 3.3 são claros quanto a direção dos efeitos dos fatores de risco modificáveis sobre o baixo desempenho escolar. As diferenças no resultado educacional entre os grupos exposto e de controle ostentam estimativas estatisticamente significativas à 1%, bem como os sinais dos coeficientes são positivos com as magnitudes praticamente invariantes as diferentes formas de pareamento. Com isso, o efeito para o cigarro é de

aproximadamente 8,5% na primeira definição, e superior a 12% na segunda. Os impactos para o consumo de bebida alcoólica e o sobrepeso seguem essa mesma tendência apontada pelos indivíduos expostos ao cigarro, só que com efeitos ligeiramente menores para a bebida alcoólica (beirando os 4% na definição 1 e 9% na definição 2) e com efeitos de magnitude ainda menores para o excesso de peso (variando de 1,6% a 2,4%).

Na comparação entre os efeitos calculados diretamente pela estimação da função de produção educacional (subseção 3.5.1) e os mensurados por PSM – mesmo presumindo diferentes controles nas duas estratégias de estimação –, o papel dos fatores de risco à saúde da criança sobre o resultado na escola permanece, inclusive quando se verifica a tendência do efeito médio (τ_k) em diferentes níveis socioeconômicos dos discentes (ver Figura 3.5). Nessa figura, desenvolvida por decis do NSE do aluno para os efeitos médios calculados para o modelo PSM-V (CR), sublinha-se que o efeito médio da exposição do fator k é negativamente correlacionado com o *background* socioeconômico.

Figura 3.5 – Efeito dos fatores de risco comportamentais à saúde sobre o atraso escolar por decis do NSE no Brasil 2012 – condicionado ao contrafactual observável



Fonte: Elaboração própria a partir dos microdados da PENSE 2012.

Nota: Efeitos computados pela abordagem PSM-V com reposição para a segunda definição de exposição ao cigarro e álcool.

Assim, as crianças mais pobres expostas aos fatores examinados são mais propensas a um menor resultado escolar. Portanto, as evidências sugerem que programas de enfrentamento a essas práticas não saudáveis poderiam beneficiar de forma mais intensa os grupos com níveis de renda mais baixos, sendo não apenas uma ação de promoção a saúde mas também uma política direcionada a justiça social, com possíveis atenuantes da desigualdade de renda. Na literatura econômica, [Conti, Heckman & Urzua \(2010\)](#) destacam a relação benéfica entre a saúde da criança, resultados na escola e desdobramentos na

trajetória interpessoal de longo prazo.

Do ponto de vista das políticas sobre o enfrentamento das DCNT no Brasil, existe uma preocupação dos *policy makers* com os fatores de risco modificáveis, dado, por exemplo, o plano de ações estratégicas para o enfrentamento das DCNT (BRASIL, 2011). Nesse cenário, os achados desta pesquisa ratificam a relevância da redução da prevalência dos fatores de risco modificáveis na população mais jovem, particularmente para os que dispõem de menor NSE. Ações de tal natureza podem resultar, inclusive, em ganhos significativos na *performance* educacional dos indivíduos.

Os achados dessa etapa da pesquisa partem de uma hipótese muito restritiva (CIA), cuja a presença de variáveis omitidas na especificação do modelo *propensity score* pode resultar em um viés na estimação dos efeitos médios sobre o grupo exposto. Segundo Rosenbaum (2010), como essa hipótese não pode ser testada diretamente, é preciso fazer pelo menos uma análise de sensibilidade desses estimadores.

3.5.3 Análise de sensibilidade: qual o tamanho do viés de fatores não-observados?

A análise de sensibilidade não exclui a possibilidade de que fatores não-observados existam, mas aponta o quão danoso eles são para os resultados. Com base nessa análise a Tabela 3.4 exibe os chamados limites de Rosenbaum para cada uma das diferentes especificações de pareamento e tipo de exposição investigado. As informações presentes na tabela exibem o primeiro valor de Γ com pelo menos 10% de significância estatística sob a hipótese de sobrestimação e sua respectiva probabilidade de não rejeição da hipótese nula (p^+).

O valor de Γ reflete a suposição sobre problemas de endogeneidade presentes nas chances de atribuição da exposição motivado por uma covariada não-observada. Para cada modelo foi calculado um intervalo de níveis críticos com suas probabilidades de não-rejeição da hipótese de sobrestimação dos resultados. Os valores presentes na tabela indicam o limite mínimo com não-rejeição da hipótese nula em que o efeito do tratamento sofre um distúrbio decorrente da autosseleção endógena do estado de exposição (ROSENBAUM, 2002).

Pela Tabela 3.4, a interferência provocada pelos fatores não-observados varia consideravelmente entre os indicativos de exposição. Os efeitos médios do excesso de peso sobre o atraso escolar são os menos robustos a possível presença de viés de seleção, já que o seu nível crítico de Γ é igual a 1,10 para todos os casos, isto é, o viés omitido capaz de gerar questionamentos desses resultados ocorre quando fatores não incluídos no modelo respondem por uma diferença de 10% na razão de chances de atribuição da exposição de excesso de peso entre os grupos de estudantes exposto e não-exposto. Para os demais fatores de risco, o nível crítico de Γ tem um valor maior, indicando uma menor

sensibilidade dessas estimativas, pois seria requerido na pior das hipóteses um viés oculto de 1,40 vezes no cigarro (1), de 1,60 no cigarro (2) e de 1,45 no álcool (2) para tornar espúria a conclusão dos efeitos sobre o indicador educacional. Para o álcool (1), o nível crítico varia de 1,10 a 1,20, conforme a técnica de pareamento adotada.

Tabela 3.4 – Análise de sensibilidade aos fatores não-observados por tipo de exposição e diferentes técnicas de pareamento nos estimadores por PSM

Modelo	Limites	Cigarro(1)	Cigarro(2)	Alcool(1)	Alcool(2)	Sobrepeso
PSM-V (CR)	Γ	1,45	1,60	1,15	1,60	1,10
	p^+	0,14	0,13	0,10	0,42	0,26
PSM-V (CR, SC)	Γ	1,45	1,60	1,15	1,60	1,10
	p^+	0,14	0,13	0,20	0,42	0,26
PSM-V (CR, Caliper)	Γ	1,45	1,60	1,15	1,60	1,10
	p^+	0,16	0,13	0,22	0,42	0,24
PSM-V (CR, Caliper, SC)	Γ	1,45	1,60	1,15	1,60	1,10
	p^+	0,16	0,13	0,22	0,42	0,24
PSM-V (SR)	Γ	1,45	1,60	1,20	1,50	1,10
	p^+	0,45	0,15	0,28	0,22	0,26
PSM-V (SR, Caliper)	Γ	1,45	1,60	1,20	1,45	1,10
	p^+	0,40	0,15	0,37	0,14	0,24
PSM-V10 (SR, Caliper)	Γ	1,45	1,75	1,10	1,70	1,10
	p^+	0,14	0,11	0,23	0,42	0,40
PSM-R	Γ	1,70	2,00	1,10	1,70	1,10
	p^+	0,16	0,23	0,25	0,31	0,48
PSM-K	Γ	1,70	1,95	1,10	1,70	1,10
	p^+	0,13	0,11	0,26	0,33	0,42

Fonte: Elaboração própria a partir dos microdados da PENSE 2012.

Nota: Γ é o primeiro valor com significância estatística do intervalo para a suposição de sobrestimação; p^+ refere-se a probabilidade de não rejeição da hipótese de sobrestimação no efeito da exposição.

Legenda: CR=com reposição; SR=sem reposição; SC=suporte comum; *Caliper*, nível de tolerância, de 0,1%. O SC impõe um suporte comum que desconsidera observações de tratamento com *propensity score* maior que o valor máximo do escore dos controles ou menor do que o valor mínimo.

No caso do cigarro, por exemplo, os fatores não-observados deveriam ser pelo menos 1,6 vezes maiores nos alunos expostos a esse fator de risco à saúde do que nos alunos não-expostos. Só assim, a probabilidade de exposição ao cigarro poderia ser considerada sobrestimada. A título de exemplo, uma variável não-observada para resultar uma mudança desse tipo deveria ter um coeficiente maior do que variáveis relativas ao contexto familiar situacional estabelecida por um pai ou por uma mãe fumante, que de acordo com o modelo referente ao cigarro (1) na Tabela 3.2 exercem uma contribuição de aproximadamente 1,4 vezes na razão de chances da criança ter experiência com o produto.

DiPrete & Gangl (2004) frisam que os limites de Rosenbaum impõem os piores cenários dos resultados, transmitindo informações sobre o grau de incerteza contido nos estimadores obtidos por *matching*. Nas estimativas deste estudo, tal análise evidencia que, até para os efeitos mais sensíveis aos fatores não-observados, existe ainda uma margem para que os resultados não possam ser enquadrados como viesados. Até porque, um Γ de 1,10 não implica que os efeitos da exposição não sejam positivos sobre o atraso escolar, mas que o intervalo de confiança para o supracitado efeito inclua zero se uma variável

não-observada provoca uma diferença em 1,10 vezes na razão de chances de atribuição da exposição entre as unidades expostas e de controle. Logo, os achados empíricos, discutidos ao longo da seção de resultados, trazem indícios críveis sobre o papel de uma postura inadequada à saúde infantil no aumento do atraso escolar.

3.6 Conclusões

Este capítulo apresenta direcionamentos empíricos sobre a relação de posturas de risco modificáveis à saúde na fase escolar e o desempenho educacional do indivíduo. Como ressaltado, esta pesquisa faz um incremento na literatura nacional, pois os trabalhos de complementariedade entre saúde e educação no país ainda não são muito explorados, bem como eles até então destacam basicamente o papel de condições antropométricas sobre o desempenho escolar com amostras menos representativas para o território nacional.

No desenvolvimento da presente pesquisa houve uma preocupação em verificar a robustez das estimativas calculadas, por isso a utilização de recortes amostrais específicos, diferentes métodos estatísticos e técnicas de pareamento e análises de sensibilidade para identificar a manutenção dos resultados para distintos grupos de controle e o nível de incerteza gerado por fatores não-observados capazes de alterar as conclusões sobre a direção dessas estimativas. Os resultados centrais corroboram a hipótese de que crianças expostas aos fatores de risco comportamentais à saúde possuem um maior atraso escolar. Sendo assim, ações que envolvam o enfrentamento das DCNT, no tocante aos fatores de risco modificáveis, implicam não apenas melhorias na saúde física e psicológica, mas também numa melhoria do rendimento escolar.

A promoção de programas educativos na escola para mostrar os malefícios desses fatores de risco à saúde e ao resultado escolar seria valioso para os estudantes, a julgar pela forte influência dos *peer-effects* no ambiente escolar sobre a probabilidade da criança decidir adotar posturas de risco, dado que altas taxas de prevalência de colegas da escola expostos repercutem diretamente na decisão individual. Ademais, políticas de promoção a saúde que afetem os hábitos da família também são relevantes, principalmente os destinados ao enfrentamento dos hábitos não saudáveis praticados em comunidades mais carentes. Pois, as diferenças no resultado escolar estimuladas por tais hábitos ocorrem com mais força entre as crianças com o NSE mais desfavorável. Nessa linha, investimentos na primeira infância podem ser uma alternativa para melhorar as condições de saúde e os rendimentos acadêmicos, já que nessa fase as habilidades cognitivas e sócio-emotivas estão em processo de formação (CONTI; HECKMAN; URZUA, 2010).

Por fim, muito embora as evidências encontradas indiquem que a exposição aos fatores de risco comportamentais à saúde afeta a *performance* na escola, as investigações futuras podem aprofundar esses achados a partir da utilização de pesquisas longitudinais,

que acompanhem a mesma criança ao longo da vida escolar, inclusive verificando seus indicadores de saúde física e psicológica, para assim estimar com maior robustez o impacto da exposição aos fatores em destaque. Além do mais, as bases de dados oficiais deveriam integrar informações sobre hábitos não saudáveis e notas em testes padronizados dos alunos, de modo a possibilitar a análise desses comportamentos sobre as habilidades cognitivas.

4 Tabagismo, penalização salarial e diferenciação compensatória para risco de acidentes no mercado de trabalho brasileiro

4.1 Introdução

As pesquisas dirigidas ao capital humano realçam a importância dos componentes educação e saúde para os retornos econômicos (SCHULTZ, 1961; MINCER, 1974; BARTEL; TAUBMAN, 1979; DEOLALIKAR, 1988). Conforme Grossman (1972), a saúde difere das outras formas de capital humano, pois o seu estoque determina a quantidade total de tempo e a intensidade que o agente se dedica às atividades produtivas e ao lazer, relacionando-se diretamente com os rendimentos individuais e com a sensação de bem-estar. Fatores não-controláveis pelo indivíduo, como herança genética, e controláveis, como hábitos saudáveis, definem o seu estoque de saúde. Logo, a adoção de comportamentos prejudiciais à saúde podem exercer um impacto negativo sobre a produtividade.

O consumo de determinadas substâncias nocivas à saúde, como aquelas reunidas no cigarro, tem ganhado destaque nesse tipo de análise, motivado por suas consequências na redução da produtividade marginal do trabalho (LEIGH; BERGER, 1989; LEVINE; GUSTAFSON; VELENCHIK, 1997; VISCUSI; HERSCH, 2001; MUNASINGHE; SICHERMAN, 2006; ANGER; KVASNICKA, 2010; BOCKERMAN; HYYTINEN; KAPRIO, 2014). Para o caso específico do cigarro, os achados empíricos da literatura internacional suportam a hipótese da existência de um custo monetário indireto do tabagismo nos ganhos individuais como, por exemplo, em Levine, Gustafson & Velenchik (1997) que estimam perdas salariais variando entre 4% e 8% para trabalhadores dos Estados Unidos. Bockerman, Hyytinen & Kaprio (2014) ressaltam ainda que o descasamento das maturidades envolvendo os custos (pecuniários e não-pecuniários, a maior parte com maior prazo de maturação) e os benefícios (sensação de prazer imediata com dependência do estoque passado) implicam em uma visão míope dos indivíduos quanto a utilização do tabaco, dado perdas salariais no longo prazo.

Nesse campo de pesquisa destaca-se também que fumantes revelam uma maior predisposição em assumir ações de risco à saúde, auferindo taxas de compensações salariais para atividades insalubres inferiores as recebidas pelos os não-fumantes. Para um mesmo *background* individual, os usuários do cigarro escolhem um ponto sobre a curva de oferta que reflete um menor salário requerido para um determinado patamar de risco, devido à subestimação acerca dos perigos de acidentes do trabalho (HERSCH; VISCUSI, 1990;

VISCUSI; HERSCH, 2001). No tocante aos resultados da modelagem sobre diferenciais compensatórios salariais para ocupações insalubres, Viscusi & Hersch (2001) mostram que fumantes estão dispostos a ganhar metade do prêmio ao risco de acidentes do trabalho estimado para os não-fumantes.

No Brasil, os dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) 2008 expõem que aproximadamente 21% de indivíduos entre 25 e 60 anos de idade são atuais usuários do cigarro, destes a maior parcela é do sexo masculino (52,1%). Dentro do grupo dos não-usuários, 18% são ex-fumantes, demonstrando que mais de 1/3 da população brasileira em idade produtiva consumiu ou consome cigarro. Considerando esse quantitativo de usuários e a escassez de pesquisas que examinam as consequências do uso do cigarro sobre indicadores do mercado de trabalho brasileiro, torna-se relevante investigar essa temática através do prisma microeconômico do retorno individual.

Para preencher essa lacuna na literatura nacional, o presente estudo examina como hábitos prejudiciais à saúde, capturada pelo tabagismo, associam-se com os resultados no mercado de trabalho por parte de agentes em fase produtiva. Mais precisamente, investiga-se a penalização salarial atribuída ao uso do cigarro e a diferenciação compensatória salarial para atividades com maiores índices de acidentes não-fatais do trabalho entre fumantes e não-fumantes. Essas análises possibilitam compreender como o uso de cigarro por parte de indivíduos na fase adulta repercute em termos de ganhos salariais e no *trade-off* envolvendo salário e risco de acidentes do trabalho. Na finalidade de responder os objetivos, estimadores quantílicos com controle para endogeneidade e viés de seleção são empregados para examinar as estimativas ao longo de diferentes pontos da distribuição condicional do salário-hora de trabalhadores brasileiros.

As principais referências que dão suporte a este ensaio são Hersch & Viscusi (1990), Levine, Gustafson & Velenchik (1997), Viscusi & Hersch (2001), Viscusi & Aldy (2003), Auld (2005), Munasinghe & Sicherman (2006), Grafova & Stafford (2009), Fernandez & Nordman (2009), Anger & Kvasnicka (2010) e Bockerman, Hyytinen & Kaprio (2014). Além do mais, os dados utilizados são provenientes da PNAD 2008 e do Anuário Estatístico de Acidentes do Trabalho (AEAT) 2008 do Ministério da Previdência Social (MPS).

Além desta introdução, a pesquisa é composta por mais cinco partes. A Seção 4.2 faz uma breve revisão da literatura sobre o tema em estudo, envolvendo os custos monetários indiretos do tabagismo no mercado de trabalho e o modelo teórico-empírico de Hersch & Viscusi (1990) e Viscusi & Hersch (2001) sobre diferenciação compensatória salarial entre fumantes e não-fumantes. A Seção 4.3 e a Seção 4.4 detalham o modelo empírico e a base de dados da pesquisa. Ao final, a Seção 4.5 e a Seção 4.6 apresentam os principais resultados e as considerações finais.

4.2 Revisão da literatura sobre tabagismo e retornos no mercado de trabalho

Tradicionalmente, como enfatizam Willis (1986) e Psacharopoulos & Patrinos (2004), as análises sobre os determinantes do retorno salarial do indivíduo se concentram no componente educação do capital humano, sendo a função de ganhos de Mincer (1974) a formulação que baliza grande parte das pesquisas econômicas na área. Estudos como Bartel & Taubman (1979), Deolalikar (1988), Hersch & Viscusi (1990), Levine, Gustafson & Velenchik (1997), Thomas & Strauss (1997), Strauss & Thomas (1998), Zarkin *et al.* (1998), Viscusi & Hersch (2001), Auld (2005), Munasinghe & Sicherman (2006), Anger & Kvasnicka (2010), Bockerman, Hyytinen & Kaprio (2014), entre outros, acrescentam o componente saúde (\mathbf{K}_i^H) na expressão de rendimentos minceriana (Equação 4.1), sob a hipótese que o estoque e comportamentos atrelados a saúde exercem efeitos na produtividade do trabalho:

$$W_i = w(\mathbf{K}_i^H, \mathbf{K}_i, \mathbf{X}_i) + \xi_i, \quad (4.1)$$

onde: W_i refere-se ao logaritmo do salário; \mathbf{K}_i^H é o vetor de medidas do componente saúde do capital humano; \mathbf{K}_i é o vetor de medidas não relativas à saúde do capital humano (como escolaridade e experiência); \mathbf{X}_i é o vetor de outras covariadas (idade, raça, gênero etc.); ξ_i é um termo de erro estocástico.

Tratando o estoque de saúde como um bem normal, Grossman (1972) assume que a saúde entra diretamente na função de preferência dos agentes (u), pois problemas de saúde geram uma redução no tempo disponível para o usufruto de atividades econômicas (como o trabalho) e não-econômicas (como o lazer). A função de utilidade intertemporal de um agente típico é expressa pelo consumo de bens de saúde (G^H) e outros bens em geral (G), tal como evidenciado na Equação 4.2

$$u = u(\phi_0 G_0^H, \dots, \phi_n G_n^H, G_0, \dots, G_n), \quad (4.2)$$

onde: G_0^H = estoque herdado de saúde; G_t^H = estoque de saúde no t -ésimo período de tempo, ϕ_t = fluxo de serviço de saúde por unidade de estoque, $\phi_t G_t^H$ representa o consumo total dos serviços de saúde; G_t = consumo total de outros bens e serviços no t -ésimo período de tempo; o indexador n é uma variável endógena que reflete a duração da vida do indivíduo, que chega ao fim quando $G_t^H = G_{min}^H$.

Nessa direção, as escolhas dos indivíduos por um estilo de vida mais saudável afetam a produção do estoque de saúde, no qual o consumo do cigarro representa uma redução no nível de investimento líquido desse estoque, $G_{t+1}^H - G_t^H$ (GROSSMAN, 1972). Como resposta da redução do investimento no capital saúde e de outras indicações comportamentais reveladas pelo consumo do produto, Levine, Gustafson & Velenchik (1997), Schwarze &

Heineck (2003), Van Ours (2004), Auld (2005), Munasinghe & Sicherman (2006), Lokshin & Beegle (2006), Grafova & Stafford (2009), Anger & Kvasnicka (2010) e Bockerman, Hyttinen & Kaprio (2014), mostram empiricamente que o tabagismo provoca uma penalização na taxa de retorno individual no mercado de trabalho. A Tabela 4.1 sumariza os resultados dos citados trabalhos.

Tabela 4.1 – Resultados sobre penalização salarial do tabagismo na literatura especializada

Estudo	Local	Amostra	Método	Período	N	Penalização (%)
Leigh & Berger (1989)	EUA	Ambos 18 ou mais	MQO	1973	1.142	-3,5 ^a
Levine <i>et al.</i> (1997)	EUA	Ambos	MQO	1984	2.817	-4,2
				1991	3.228	-6,9
			Diferenças entre irmãos	1984	151	-8,0
			Pooling	1991	228	-8,1
Schwarze & Heineck (2003)	ALE	Homens 25-55 anos	MQ2E	1998	2.570	-2,5
			Painel	1998-2001	8.595	-0,7 ^a
Van Ours (2004)	HOL	Homens 26-55 anos	MQO	2001	503	-10,7
			MQ2E	2001	503	-16,6
Auld (2005)	CAN	Homens 25-59 anos	MV	1985 e 1991	3.891	-8,3
			FIMSL-VI	1985 e 1991	3.891	-24,0
Munasinghe & Sicherman (2006)	EUA	Ambos	Dinâmico(1)	1979-1994	9.396	-0,5
			Dinâmico(2)	1979-1994	9.396	-0,2
			Dinâmico(3)	1979-1994	9.396	-0,2
Lokshin & Beegle (2006)	ALB	Homens 25-60 anos	MQO	2005	1.588	-4,8
			MQ2E	2005	1.588	-25,6
Braakmann (2008)	RU	Homens	MQO	1991-2005	32.356	-3,1
			Painel-EF	1991-2005	32.356	-0,9 ^a
			Painel-VI	1991-2005	17.216	0,2 ^a
Grafova & Stafford (2009)	EUA	Homens 25-60 anos	MQO	1986	3.056	-3,4
				1999	3.062	-9,2
				2001	2.898	-10,9
			MQO-pooled	1986-2001	9.016	-7,4
			MQO-EF	1986-2001	9.016	0,9
Anger & Kvasnicka (2010)	ALE	Homens 27-55 anos	MQO	2002	1.968	-4,5
			MQ2E	2002	1.968	-9,9
Bockerman <i>et al.</i> (2014)*	FIN	Ambos	MQO	1990-2004	3.914	-2,16
			Gêmeos	1990-2004	1.957	-1,41
			Gêmeos-DZ	1990-2004	1.350	-1,30
			Gêmeos-MZ	1990-2004	607	-1,85

Fonte: Elaboração própria.

Legenda (Local): EUA = Estados Unidos; ALE = Alemanha; HOL = Holanda; CAN = Canadá; ALB = Albânia; RU = Reino Unido; FIN = Finlândia.

Legenda (Método): MQO = Mínimos Quadrados Ordinários; MQ2E = Mínimos Quadrados em dois estágios; MV = Máxima Verossimilhança; FIMSL = *full information maximum simulated likelihood*; Gêmeos = modelo de diferenças entre gêmeos; DZ = gêmeos dizigóticos; MZ = gêmeos monozigóticos. Em Munasinghe & Sicherman (2006), Dinâmico(1), Dinâmico(2) e Dinâmico(3) representam a regressão sem controle, com controles limitados e controles completos.

^a = Não significativo estatisticamente; * Variável explicativa é o número de pacotes de cigarro consumidos.

Agrupando os estudos por país ou região, tem-se inicialmente que Leigh & Berger (1989) são um dos primeiros a averiguar a relação entre tabagismo e salário, encontrando uma associação sem significância estatística. Neste estudo, os dados do *Quality of Employment Survey* (QES) de 1973 para trabalhadores dos Estados Unidos (EUA) são usados em conjunto com modelo clássico de regressão, no qual além do impacto do uso do cigarro, avalia-se também o papel do excesso de peso sobre os salários – que também não apresenta

significância. Já em [Levine, Gustafson & Velenchik \(1997\)](#), os fumantes, a partir dos dados do *National Longitudinal Survey of Youth* (NLSY) de 1984 e 1991 para os EUA, têm uma diminuição nos salários entre 4% e 8%, dependendo da especificação do modelo, quando comparado aos não-fumantes. Os autores adotam diferentes estratégias, dentre elas a utilização do modelo de efeito fixo do componente familiar na regressão de diferenças salariais entre os irmãos, que visa controlar características não observáveis da família. Além deles, [Munasinghe & Sicherman \(2006\)](#) e [Grafova & Stafford \(2009\)](#) desenvolvem aplicações para trabalhadores norte-americanos, dando suporte aos efeitos negativos do cigarro sobre a taxa de salário.

[Grafova & Stafford \(2009\)](#) examinam os diferentes grupos de fumantes em termos de ganhos salariais, concentrando-se nos usuários persistentes que consomem o cigarro por um longo período de tempo. Com base nos dados do *Panel Study of Income Dynamics* (PSID) nos anos de 1986, 1999 e 2001, os mesmos aferem diferenças salariais significativas entre fumantes e não-fumantes. Por MQO agrupado, os fumantes recebem, em média, 7,4% a menos que os não-fumantes. Considerando o histórico dos indivíduos, os consumidores persistentes quando confrontados com diferentes bases de especificação (indivíduos que nunca fumaram e ex-fumantes) apresentam variações salariais estatisticamente significativas variando de -8% a -12%, dependendo do período.

O trabalho de [Auld \(2005\)](#) tem por escopo aferir de forma conjunta o papel do consumo de bebidas alcoólicas e do cigarro sobre os salários de indivíduos canadenses a partir do *General Social Survey* (GSS) para os anos de 1985 e 1991. As suas principais descobertas dizem respeito à subestimação da penalização salarial do tabagismo, na abordagem sem controle para endogeneidade, em cerca de 1/3. Comparativamente, os efeitos do cigarro se mostram bem maiores do que o de consumo de bebidas alcoólicas sobre os salários, que em alguns casos os parâmetros referentes ao álcool não são diferentes de zero na perspectiva estatística.

Em relação aos trabalhos com dados europeus, [Schwarze & Heineck \(2003\)](#), através do *Socio-Economic Panel* (SOEP) para Alemanha, obtêm na abordagem com dados de corte que os fumantes do sexo masculino têm uma redução na taxa de salário de aproximadamente 2,5%, enquanto que no modelo com efeito fixo, controlando a heterogeneidade individual não-observada, o impacto do cigarro sobre os salários não é significativo estatisticamente. Já [Anger & Kvasnicka \(2010\)](#), com a mesma base de dados, acham efeitos negativos do consumo do tabaco sobre os ganhos dos indivíduos, sendo os custos indiretos do cigarro na abordagem com variáveis instrumentais (VI) representado por uma redução em 0,099 do log do salário-hora.

Fazendo a varredura de estudos para outros países europeus, apenas o trabalho de [Braakmann \(2008\)](#), para dados do *British Household Panel Survey* (BHPS) de 1991 a 2005, não suporta a hipótese que relaciona a penalização salarial de fumar e menor

produtividade dos fumantes nos modelos de efeito fixo e VI. Muito embora, identifica uma penalização de 3% com dados de corte. As demais pesquisas, como [Van Ours \(2004\)](#) para Holanda, [Lokshin & Beegle \(2006\)](#) para Albânia e [Bockerman, Hyytinen & Kaprio \(2014\)](#) para Finlândia, corroboram os achados da grande maioria da literatura sobre a diminuição nos rendimentos provocada pelo tabagismo.

Dentre as localidades reportadas na Tabela 4.1, a que possui o menor nível de renda *per capita* é a Albânia, onde [Lokshin & Beegle \(2006\)](#) fazendo uso do *Living Standards Monitoring Survey* (LSMS) de 2005, constatam uma penalização salarial do tabagismo superior a 20% no modelo com VI. Usando as pesquisas *Older Finnish Twin Cohort Study* (OFTCS) do Departamento de Saúde Pública da Universidade de Helsinque na Finlândia e *Finnish Longitudinal Employer-Employee Data* (FLEED) para gêmeos do sexo masculino nascidos entre 1945 e 1957, [Bockerman, Hyytinen & Kaprio \(2014\)](#) concluem que fumantes possuem menor rendimento de longo prazo, controlando por fatores ambientais, efeitos dos pares e carga genética (ao usar dados sobre gêmeos idênticos, monozigóticos).

Quanto aos motivos discutidos na literatura que podem explicar a penalização salarial dos fumantes no mercado de trabalho, enumeram-se os seguintes:

- Interferência do consumo do cigarro em tarefas de execução manual;
- Elevação dos custos relativos do empregador, devido, por exemplo, ao maior absenteísmo e custos do seguro-saúde;
- Discriminação no ambiente de trabalho causada por questões de atratividade física visual (possíveis danos na pele) ou olfativa (decorrente do cheiro de fumaça);
- Predisposição dos fumantes por emprego que disponibilizam seguro-saúde em detrimento de uma maior remuneração;
- Alta taxa de desconto intertemporal, indicando maior propensão ao consumo presente em relação ao consumo futuro e, assim, baixos investimentos no capital humano. ([LEVINE; GUSTAFSON; VELENCHIK, 1997](#); [GRAFOVA; STAFFORD, 2009](#)).

A literatura empírica sobre essa temática chama atenção, em especial, para a influência do último ponto acima na estimativa da penalização salarial, haja vista o valor não-observado dessa taxa de desconto que pode influir simultaneamente na decisão de fumar e na taxa de salário. Neste quadro, se uma pessoa com preferências imediatistas tem mais predisposição a fumar, o parâmetro do custo indireto do tabagismo seria tendencioso, porque as consequências negativas sobre os salários de tais preferências pelo consumo presente, dado o menor investimento no capital humano, seriam atribuídas ao ato de fumar. Assim, o indicador do status de fumante (H_i) é tratado como um regressor endógeno,

$Cov(H_i, \epsilon_i) \neq 0$, o que demanda um tratamento para o problema de endogeneidade. Logo, a grande maioria dos estudos aplicados na área tenta contornar a questão dos fatores não-observados, como fica demonstrado pelo crescente rigor metodológico das pesquisas mencionadas.

Muito embora as estimativas das penalizações salariais do tabagismo possam chegar em alguns casos até a 25,6% na Tabela 4.1, ainda não existem aferições dos efeitos do uso do cigarro sobre os resultados do mercado de trabalho brasileiro⁴¹. De uma forma global, a exploração do assunto está concentrada nos países da América do Norte e da Europa sem a incorporação da heterogeneidade da penalização salarial do uso do cigarro para trabalhadores, por exemplo, com baixos e altos níveis de produtividade do trabalho. Assim, considerando que a posição na distribuição dos rendimentos carrega certas peculiaridades do tipo de função desempenhada pelo trabalhador – por exemplo, ocupações com altas remunerações são mais propensos a requerer mais conhecimento técnico do que condicionamento físico –, o tabagismo pode ter diferentes repercussões ao longo da função quantílica condicional dos salários.

4.2.1 Diferenciação compensatória salarial para trabalhadores fumantes

Para Smith (1979), a taxa de salário contém uma série de elementos relevantes para a tomada de decisão de ambos os lados da oferta e demanda do mercado de trabalho, a saber: intensidade do trabalho, probabilidade de acidente, tarefas insalubres ou que geram desutilidade etc. Sob os fundamentos da teoria dos salários hedônicos⁴², a utilidade do trabalhador é crescente com a taxa de salário (w) e decrescente com os indicadores de risco de acidentes do trabalho fatais (r) e não-fatais (p). Com efeito, w incorpora os preços implícitos desses riscos atrelados ao exercício do emprego, denominados de diferenciais compensatórios, $w'(p) = \partial w / \partial p$ ou $w'(r) = \partial w / \partial r$, que correspondem a inclinação da curva do salário hedônico.

Tratando do risco de acidentes não-fatais do trabalho, $w'(p)$ representa na margem o *value of a statistical injury*⁴³ (VSI), enquanto o *value of a statistical life* (VSL) é específico para os casos de riscos de acidentes fatais, $w'(r)$. Essas medidas são comumente utilizadas na área para expressar o pagamento adicional que os trabalhadores esperam pela periculosidade do trabalho. Viscusi & Aldy (2003), que fazem um *overview* dessa literatura nos últimos 30 anos, confirmam a hipótese de diferenciais salariais para compensar os riscos ocupacionais para diferentes tipos de modelagens econométricas e para uma série

⁴¹ Grande parte da literatura nacional, balizada pela Equação 4.1, está voltada para o papel das condições nutricionais/antropométricas nesse processo, como é o caso de Thomas & Strauss (1997) e Kassouf (1997), ou faz uso de classificações do estado geral de saúde, como Alves & Andrade (2003).

⁴² A teoria dos salários hedônicos presume conceitos estabelecidos pela teoria dos preços hedônicos apresentado por Rosen (1974).

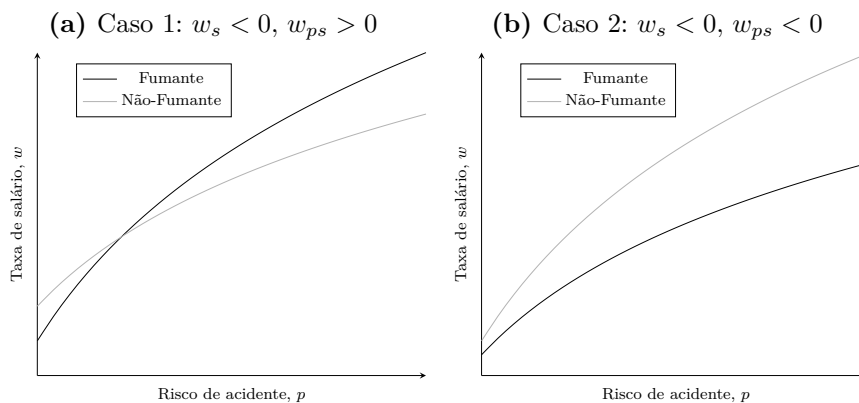
⁴³ Valor estatístico de um acidente é a tradução desse termo adotada nesta pesquisa.

países⁴⁴. Os autores identificam ao longo dos diversos trabalhos empíricos para os EUA que, na mediana, o VSL anual equivale a US\$ 7 milhões de dólares, enquanto que em países em desenvolvimento o VSL é bem inferior do que o estimado nas localidades mais desenvolvidas, sinalizando uma maior sensibilidade das curvas de salário hedônico em economias com maior nível de renda *per capita*. Quanto ao VSI, os valores estimados, a preços constantes do ano 2000, para o mercado de trabalho dos EUA se situam entre US\$ 20 mil e US\$ 70 mil.

Nesse contexto, o trabalho de [Hersch & Viscusi \(1990\)](#) explora até que ponto a adoção de hábitos pessoais que revelam atitudes ao risco (como o uso do cinto de segurança e de cigarro) podem refletir no *trade-off* salário-risco. Os achados para trabalhadores norte-americanos explicitam que os hábitos pessoais relacionados à saúde determinam a taxa de compensação requerida para atividades arriscadas, com os fumantes tendendo a subestimar os riscos (recebendo, em média, uma menor taxa de compensação salarial por unidade de risco do trabalho) e os indivíduos que revelam menor aversão ao risco requerendo uma maior compensação para atividades mais arriscadas.

[Viscusi & Hersch \(2001\)](#) desenvolvem um modelo alicerçado nas hipóteses de [Hersch & Viscusi \(1990\)](#), em que as preferências pessoais de risco à saúde dos trabalhadores são associadas com a taxa de salário de compensação e a escolha ocupacional dos agentes. Trabalhadores fumantes e não-fumantes são divididos em grupos com preferências ao risco distintas quando se considera o *loci* salário-risco (como exibido na Figura 4.1). Destarte, os usuários do cigarro revelam atitudes e predisposição ao risco, pois, por exemplo, a exposição ao cigarro gera uma menor expectativa de vida, bem como sinaliza uma maior taxa de preferência temporal com respeito aos anos de vida ([HERSCH; VISCUSI, 1990](#); [VISCUSI; HERSCH, 2001](#); [SCHARFF; VISCUSI, 2011](#)).

Figura 4.1 – Relação salário-risco por status de fumante para dois casos hipotéticos



⁴⁴ [Esteves \(2008\)](#), usando dados de 1997 a 1999, direcionados para trabalhadores formais, da Relação Anual de Informações Sociais (RAIS) e do AEAT/MPS, explicita a existência de diferenciais compensatórios para o Brasil.

O modelo sobre tabagismo e diferenciação compensatória incorpora o papel do status de fumante nos dois lados do mercado, com o escopo de avaliar o papel das preferências de risco usando a condição de fumante como um indicador de atitudes em face ao risco. A taxa de salário, o prêmio ao risco e os riscos de acidentes do trabalho são denotados, respectivamente, por w , w_p e p , com $0 \leq p \leq 1$ e $w_p > 0$, pois existe um custo marginal positivo para a segurança do trabalho. Admite-se dois tipos de função utilidade: $U_1[w(p, s)]$, que reflete a utilidade para boas condições de saúde; $U_2[w(p, s), h(s)]$, para o caso da utilidade em condições de acidentado. Nesse caso, s representa a intensidade em que o indivíduo utiliza cigarro (se $s = 0$, tem-se que o indivíduo é não-fumante) e $h(s)$ é um índice de utilidade não observável, que capta uma maior disposição para assumir riscos à saúde. Desse modo, o problema de otimização é estabelecido pela Equação 4.3.

$$\max_p V = (1 - p)U_1[w(p, s)] + pU_2[w(p, s)h(s)] \quad (4.3)$$

Pela condição de primeira ordem ($\frac{\partial V}{\partial p} = 0$), obtém-se o valor referente ao diferencial compensatório $w_p = \frac{\partial w(p, s)}{\partial p}$ exposto na Equação 4.4.

$$w_p = \frac{U_1[w(p, s)] - U_2[w(p, s)h(s)]}{(1 - p)\frac{\partial U_1[w(p, s)]}{\partial w(p, s)} + p\frac{\partial U_2[w(p, s)h(s)]}{\partial w(p, s)}} \quad (4.4)$$

Pela Equação 4.4, as escolhas dos agentes dependem da estrutura das funções de utilidade e das curvas de ofertas de salários. Para medir o efeito da intensidade do uso do cigarro no ponto ótimo do trabalho de risco selecionado, faz-se a diferenciação total da Equação 4.4, resolvendo para $\frac{\partial p}{\partial s}$,

$$\begin{aligned} \frac{\partial p}{\partial s} = & [-(U_{1w} - U_{2w})w_s + (1 - p)U_{1ww}w_s w_p + (1 - p)U_{1w}w_{ps} \\ & + U_{2h}h_s + pU_{2ww}w_s w_p + pU_{2wh}w_p h_s + pU_{2w}w_{ps}] \times (-D)^{-1}, \end{aligned} \quad (4.5)$$

onde: $D = -2w_p(U_{1w} - U_{2w} + (1 - p)U_{1ww}(w_p)^2 + (1 - p)U_{1w}w_{pp} + pU_{2w}w_{pp}) < 0$.

O sinal do efeito do cigarro no *trade-off* salário-risco é ditado, em especial, pelo sinal de w_{ps} . Se $w_{ps} \geq 0$, o salário de risco amplia com a intensidade do uso do cigarro, caso contrário $\frac{\partial p}{\partial s}$ pode ser negativo, desde que o efeito seja dominante. Dessa forma, Viscusi & Hersch (2001) destacam diferentes possibilidades envolvendo o efeito do tabagismo na diferenciação compensatória. Duas dessas possibilidades estão descritas na Figura 4.1

As evidências empíricas de Viscusi & Hersch (2001), usando dados do *National Medical Expenditure Survey* (NMES) de 1987 e do *Bureau of Labor Statistics* (BLS) para uma amostra de 4.821 trabalhadores do sexo masculino entre 18 e 65 anos de idade dos EUA, sustentam a suposição que fumantes são mais propensos a selecionar trabalhos

com maior probabilidade de acidentes e a receber um salário de compensação para o risco inferior do que os recebidos pelos não-fumantes. Conforme os valores implícitos por acidente, os fumantes recebem um valor esperado 56% menor do que o recebido pelos não-fumantes.

Acerca dos problemas empíricos de estimação do prêmio ao risco, [Garen \(1988\)](#) adverte que a endogeneidade do risco de acidentes causa viés nas estimativas, com a subestimação dos diferenciais compensatórios para riscos fatais e não-fatais por MQO. Além de [Garen \(1988\)](#), um conjunto de pesquisas, como [Viscusi & Hersch \(2001\)](#), [Viscusi & Aldy \(2003\)](#), [Esteves \(2008\)](#), [Polat \(2013\)](#) e [Contzen, Won & Lavin \(2013\)](#), sublinham o problema de endogeneidade mencionado, com a maioria adotando a abordagem de variáveis instrumentais para superá-lo. Basicamente, existem dois canais para o problema de endogeneidade: simultaneidade, visto que segurança é um bem normal, trabalhadores com mais riqueza tendem a optar por trabalhos mais seguros; existência de fatores não-observados que exercem influência tanto na percepção de risco quanto na taxa de salário. [Viscusi & Aldy \(2003\)](#) relata que o tratamento ideal dos fatores não observáveis, resultantes de diferenças nas características do trabalhador, seria por meio de dados longitudinais. Para tanto o próprio autor admite a dificuldade desse tipo de investigação pela escassez de *surveys* que seguem um conjunto de indivíduos ao longo de vários anos.

4.3 Estratégia empírica

A estratégia empírica divide-se em duas partes. A primeira descreve o modelo de penalização salarial supostamente causado pelo o uso do cigarro ao longo dos quantis condicionais dos salários. Já a segunda parte refere-se ao modelo empírico sobre a diferenciação compensatória salarial para atividades arriscadas entre fumantes e não-fumantes no mercado de trabalho.

4.3.1 Modelo econométrico para penalização salarial

Esta parte baseia-se nos trabalhos de [Levine, Gustafson & Velenchik \(1997\)](#), [Schwarze & Heineck \(2003\)](#), [Van Ours \(2004\)](#), [Auld \(2005\)](#), [Lokshin & Beegle \(2006\)](#), [Anger & Kvasnicka \(2010\)](#) e [Bockerman, Hyttinen & Kaprio \(2014\)](#). A abordagem mais recorrente na área é a de variáveis instrumentais (VI), que mensura o efeito do consumo do cigarro para um agente representativo sobre o valor condicional médio dos ganhos salariais, tratando essa variável explicativa como um regressor endógeno.

A avaliação do papel das condições de saúde e, em particular, dos hábitos não saudáveis sobre o retorno salarial possui problemas inerentes de mensuração do efeito causal em bases de dados, particularmente, sem caráter longitudinal ([LEVINE; GUSTAFSON; VELENCHIK, 1997](#); [BOCKERMAN; HYYTINEN; KAPRIO, 2014](#)). A dificuldade se deve

a atributos não observáveis dos indivíduos que afetam concomitante a decisão de fumar e o salário-hora. Assim, as estimativas obtidas pelos métodos que supõem exogeneidade, como MQO e regressão quantílica (RQ), são tendenciosas.

Neste trabalho, estimadores quantílicos com regressores endógenos são utilizados para estimar a relação entre o uso do cigarro e ganhos salariais para os n -ésimos quantis (τ) da distribuição condicional do salário-hora, onde $\tau \in (0, 1)$. A função de regressão quantílica condicional com variáveis instrumentais (RQVI), baseada em Chernozhukov & Hansen (2004, 2005, 2013) e Chernozhukov, Hansen & Jansson (2007), é a abordagem escolhida para o cômputo das estimativas de interesse em contraponto à função condicional média com variáveis instrumentais, usada em análises de regressão linear como no método de mínimos quadrados em dois estágios (MQ2E). Além de verificar a heterogeneidade dos coeficientes ao longo da distribuição condicional, a regressão quantílica é mais robusta à presença de *outliers* e de erros com distribuição não-gaussiana do que os estimadores de mínimos quadrados (KOENKER, 2005). É válido salientar que na seção de resultados desta pesquisa, as estimativas inicialmente são calculadas pelos métodos MQO, MQ2E e RQ convencional (sem controle para VI), para em seguida apresentar os coeficientes obtidos por RQVI.

Os estimadores quantílicos, como destacado por Koenker & Bassett (1978), identificam o efeito das covariadas não apenas para a tendência central, mas também ao longo da distribuição condicional da variável de *outcome*. Com base na RQ convencional, a Equação 4.1 de rendimentos do indivíduo pode ser representada pela seguinte representação linear

$$W_i = \delta(\tau)H_i + \mathbf{X}\mathbf{1}_i\boldsymbol{\beta}(\tau) + \epsilon_{1i}(\tau), \text{ com } Q_\tau(W_i|H_i, \mathbf{X}\mathbf{1}_i) = \delta(\tau)H_i + \mathbf{X}\mathbf{1}_i\boldsymbol{\beta}(\tau), \quad (4.6)$$

onde: W_i é o logaritmo do salário-hora do i -ésimo indivíduo; H_i é uma variável binária referente à condição de fumante do trabalhador; $\mathbf{X}\mathbf{1}$ é um vetor formado por variáveis de controle, o que inclui idade, faixa de escolaridade, raça, fatores locais, outros atributos pessoais e o intercepto do modelo.

O τ -ésimo quantil condicional do logaritmo do salário-hora em relação às variáveis explicativas é denotado por $Q_\tau(W_i|H_i, \mathbf{X}\mathbf{1}_i)$. Isto posto, a Equação 4.7 refere-se ao problema de minimização para a estimação dos parâmetros para um determinado quantil com $\theta = f(\delta, \boldsymbol{\beta})$:

$$\min_{\theta \in R^k} \left[\sum_{i: W_i \geq Q_\tau(\cdot)} \tau |W_i - \delta_\tau H_i - \mathbf{X}\mathbf{1}_i \boldsymbol{\beta}_\tau| + \sum_{i: W_i < Q_\tau(\cdot)} (1 - \tau) |W_i - \delta_\tau H_i - \mathbf{X}\mathbf{1}_i \boldsymbol{\beta}_\tau| \right]. \quad (4.7)$$

Nesse caso em estudo, ao averiguar o papel do regressor endógeno (H_i), o coeficiente $\delta(\tau)$ obtido a partir da regressão quantílica convencional se torna tendencioso, necessitando

adotar alguma abordagem, como variáveis instrumentais, que permita tratar esse viés. Contudo, como apresentado em Chernozhukov & Hansen (2005), as técnicas padrões de variáveis instrumentais (como MQ2E) não são consistentes com o efeito de tratamento quantílico variando ao longo dos diferentes quantis da distribuição da variável dependente. Assim, no modelo RQVI, inicialmente, a relação entre o regressor binário endógeno $H_i \in \{0, 1\}$, que representa o status de fumante do trabalhador, e o salário-hora (W_i) pode ser representada pela Equação 4.8

$$W_i = \delta(\epsilon_{1i}^*)H_i + \mathbf{X1}_i\beta(\epsilon_{1i}^*), \text{ com } \epsilon_{1i}^*|H_i, \mathbf{X1}_i \sim \text{uniforme}(0, 1), \quad (4.8)$$

onde: ϵ_{1i}^* é uma variável aleatória não-observada, representando alguma habilidade inata ou nível de preferência dos indivíduos, responsável pela heterogeneidade dos *outcomes* para os indivíduos com idênticas características observáveis ($\mathbf{X1}$) e do tratamento (D); $\tau \rightarrow \delta(\tau)H_i + \mathbf{X1}\beta(\tau)$ é a função quantil condicional de W_i . Por construção, $\delta(\tau)H_i + \mathbf{X1}\beta(\tau)$ é o τ -ésimo quantil do W_i condicionado as variáveis explicativas.

Como H_i é correlacionado com ϵ_{1i}^* , a Equação 4.8 é reescrita para responder o problema de endogeneidade, onde V_i é uma variável não-observada estatisticamente dependente do termo de erro ϵ_{1i} e possui outros fatores omitidos que afetam a decisão de fumar,

$$W_i = \delta(\epsilon_{1i})H_i + \mathbf{X1}_i\beta(\epsilon_{1i}), \text{ com } \epsilon_{1i}|H_i, \mathbf{X1}_i, \mathbf{Z1}_i \sim \text{uniforme}(0, 1) \quad (4.9)$$

$$H_i = f(\mathbf{X1}_i, \mathbf{Z1}_i, V_i) \quad (4.10)$$

$$\tau \rightarrow \delta(\tau)H_i + \mathbf{X1}_i\beta(\tau), \quad (4.11)$$

em que essa decisão é determinada pela Equação 4.10. O vetor V_i é composto por variáveis não-observadas correlacionadas com ϵ_{1i} , e $\mathbf{Z1}_i$ é um vetor de variáveis instrumentais independentes a ϵ_{1i} mas correlacionado com a decisão de fumar (H_i). As Equações 4.9 e 4.11, que balizam as condições da abordagem RQVI, implicam:

$$\Pr(W_i \leq \delta(\tau)H_i + \mathbf{X1}_i\beta(\tau)|\mathbf{X1}_i, \mathbf{Z1}_i) = \tau. \quad (4.12)$$

A Equação 4.12, segundo Chernozhukov & Hansen (2004), estabelece a conexão entre os parâmetros da RQVI para o conjunto de equações de momentos condicionais necessárias para a estimação e inferência. Para resolver a citada expressão é preciso achar a solução para regressão quantílica de $W_i - [\delta(\tau)H_i + \mathbf{X1}_i\beta(\tau)]$ em $(\mathbf{X1}, \mathbf{Z1})$, que equivale a:

$$Q(W_i \leq \delta(\tau)H_i + \mathbf{X1}_i\beta(\tau)|\mathbf{X1}_i, \mathbf{Z1}_i) = \mathbf{X1}_i\beta(\tau) + \mathbf{Z1}_i\gamma(\tau). \quad (4.13)$$

Dado o verdadeiro valor de $\delta(\tau)$, estima-se os demais parâmetros, $\beta(\tau)$ e $\gamma(\tau)$ por meio do seguinte problema de minimização:

$$\min_{\theta \in R^k} \sum_i \rho_\tau [W_i - \delta(\tau)H_i - \mathbf{X}\mathbf{1}_i\beta(\tau) - \mathbf{Z}\mathbf{1}_i\gamma(\tau)], \quad (4.14)$$

onde: ρ_τ é uma função definida como $\rho_\tau(\psi) = \tau\psi$ se $\psi \geq 0$ ou $\rho_\tau = (\tau - 1)\psi$ se $\psi < 0$, com $\psi = [W_i - \delta(\tau)H_i - \mathbf{X}\mathbf{1}_i\beta(\tau) - \mathbf{Z}\mathbf{1}_i\gamma(\tau)]$. Como não se conhece o valor de $\delta(\tau)$, o problema exposto na Equação 4.14 é solucionado pela aproximação do valor de $\gamma(\tau)$ em torno de zero.

Segundo Van Ours (2004), as variáveis instrumentais frequentemente utilizadas para resolver o problema de endogeneidade da variável explanatória fumante são formadas, em geral, por fatores atrelados à religião, ao preço do produto, a auto-avaliação e/ou ao comportamento da família. Dessa maneira, a validade de uma estimativa por VI depende do pressuposto de que os instrumentos não estão correlacionados com outras características latentes dos agentes que podem influir nos seus ganhos, não sendo portanto um exercício trivial encontrar bons instrumentos. Contudo, Auld (2005) ressalta que o preço do cigarro exerce influência na decisão de consumo do produto mas não se relaciona com a taxa de salário, sendo uma variável exógena consolidada nessa área. Além do mais, para Cutler & Glaeser (2005) e Anger & Kvasnicka (2010), existem evidências na literatura sobre o papel da percepção dos indivíduos quanto ao malefício do cigarro e as características do ambiente familiar sobre o tabagismo no hábito de fumar, onde inexistente ou é pouco provável um canal direto dessas covariadas sobre os rendimentos atuais do indivíduo.

Com base nas variáveis instrumentais adotadas por Van Ours (2004), Auld (2005) e Anger & Kvasnicka (2010), o vetor $\mathbf{Z}\mathbf{1}_i$ é composto pelo preço do cigarro, *dummy* para o indivíduo que co-reside com ao menos um ex-fumante, quantidade de pessoas fumantes na família (exclusive o próprio indivíduo), indicadores sobre crença/percepção dos agentes quanto aos riscos do uso cigarro para si e para terceiros e regras sobre o uso do cigarro no domicílio. Essas percepções, regras, preços do produto e características de outros moradores do domicílio são potenciais instrumentos por se relacionarem com a decisão de se fumar e serem condicionalmente não correlacionadas com os salários⁴⁵. É válido acrescentar que as variáveis co-residentes, conforme Anger & Kvasnicka (2010), são usadas com a finalidade de medir possíveis influências sobre o comportamento do fumante corrente, tanto em termos das chances de sua decisão de fumar ter sido definida pela existência de outros familiares fumantes ou de pessoas na família com histórico bem-sucedido de cessação do tabagismo.

Ressalta-se que a estimação da Equação 4.9, seguindo os procedimentos de Buchinsky (1998, 2001) para modelos com regressores quantílicos, é realizada em dois estágios

⁴⁵ A fim de aferir a força dos instrumentos e a restrição de sobre-identificação, a Tabela 4.5 reporta testes estatísticos para o ajustamento dos instrumentos e exogeneidade do regressor.

a fim de corrigir o problema de seleção amostral discutido em Heckman (1979), pois o termo de viés de autoseleção na força de trabalho tem uma forma desconhecida na RQ. Destarte, admite-se que o salário observado (S) pode ser escrito como

$$S_i = d_i \times W_i, \text{ com } d_i = I(W_i > W_i^R), \quad (4.15)$$

onde $I(\cdot)$ é uma função indicadora usual e W_i só é observado se exceder o salário de reserva (W_i^R), de modo que $d_i = 1$, caso o salário seja observável, e $d = 0$, caso contrário. Na presença do mecanismo de seleção, o quantil condicional do salário observado é expresso pela Equação 4.16

$$Q_\tau(S_i|H_i, \mathbf{X1}_i, \mathbf{Z1}_i) = \delta(\tau)H_i + \mathbf{X1}_i\beta(\tau) + \mathbf{Z1}_i\gamma(\tau) + h_\tau(g_0), \quad (4.16)$$

em que $h_\tau(g_0) = Q_\tau(\epsilon_{1i}|H_i, \mathbf{X1}_i, \mathbf{Z1}_i, d_i = 1)$ e $g_0 = g(\mathbf{J}_i, \boldsymbol{\kappa}_0)$. Adianta-se que o vetor de variáveis explicativas, \mathbf{J} , é formado por covariadas sobre a decisão de oferta do trabalho (com $\mathbf{X1} \subset \mathbf{J}$), incluindo o rendimento oriundo do não-trabalho (aluguel, juros, transferências, doações etc.), estrutura familiar e dificuldades físicas para realização de trabalhos pesados⁴⁶. Segundo Buchinsky (1998), a equação de participação na força de trabalho deve ser estimada por um modelo semiparamétrico. Por isso foi escolhido o modelo de resposta binária de estimadores de máxima verossimilhança semi não-paramétricos⁴⁷ (SNP) de Gallant & Nychka (1987).

Em seguida, a exemplo de Buchinsky (1998), a aproximação da razão inversa de Mills com dois termos é adicionada na equação dos salários observados dos indivíduos⁴⁸. Nesse ponto, o termo $h_\tau(g) = \varrho'_\tau P_{3j}(g)$ controla a seleção amostral e $P_{3j}(g)$ é um polinômio de terceira ordem em g , cujo $h_j(g) = \lambda(\hat{\mu} + \hat{\sigma}g)^{j-1}$, com $j = 1, 2$ e $\lambda(\cdot) = \phi(\cdot)/\Phi(\cdot)$, que é a razão inversa de Mills. $\phi(\cdot)$ e $\Phi(\cdot)$ são respectivamente a função de densidade de probabilidade (fdp) e a função de distribuição acumulada (fda), bem como $\hat{\mu}$ e $\hat{\sigma}$ são parâmetros de localização e escala.

4.3.2 Modelo econométrico: diferenciação compensatória salarial por status de fumante

O modelo empírico descrito nesta parte toma por referência Garen (1988), Hersch & Viscusi (1990), Viscusi & Hersch (2001), Fernandez & Nordman (2009) e Contzen, Won & Lavin (2013). A Equação 4.17, que capta o equilíbrio resultante do *trade-off* salário-risco

⁴⁶ A Tabela C.3 no Apêndice descreve as variáveis utilizadas para a equação de seleção.

⁴⁷ O comando semi-nonparametric (SNP) desenvolvido por De Luca (2008) é usado para a estimação desses coeficientes, supondo uma expansão polinomial univariada de terceira ordem de Hermite para fazer a aproximação da densidade desconhecida dos erros da regressão latente.

⁴⁸ No caso dos modelos de referência para a função condicional média, o cálculo da razão inversa de Mills tradicional (λ) usa o modelo probit para estimar a equação de participação.

comumente usada nessa literatura, [Smith \(1979\)](#), [Arnould & Nichols \(1983\)](#), [Viscusi & Hersch \(2001\)](#) e [Fernandez & Nordman \(2009\)](#), é o ponto de partida para avaliação dos diferenciais compensatórios salariais. Nesta expressão, inclui-se como covariada o indicador do nível de risco inerente ao desempenho de uma determinada atividade econômica⁴⁹, com regressões separadas entre fumantes (f) e não-fumantes (nf) como em [Viscusi & Hersch \(2001\)](#).

$$\begin{cases} W_{ij}^f = a_0^f + a_1^f p_{ij}^f + \mathbf{X}\mathbf{2}_i^f \boldsymbol{\alpha}^f + \epsilon_{ij}^f \\ W_{ij}^{nf} = a_0^{nf} + a_1^{nf} p_{ij}^{nf} + \mathbf{X}\mathbf{2}_i^{nf} \boldsymbol{\alpha}^{nf} + \epsilon_{ij}^{nf} \end{cases} \quad (4.17)$$

Onde: p_{ij} é o índice de risco de acidentes do trabalho para o i -ésimo trabalhador empregado na divisão econômica j . Como adotado, por exemplo, em [Viscusi & Hersch \(2001\)](#), [Esteves \(2008\)](#), [Evans & Schaur \(2010\)](#) e [Contzen, Won & Lavin \(2013\)](#), assume-se um efeito fixo dos indicadores de risco setorial de acidentes.

Para mensurar o prêmio ao risco (w_p) entre fumantes e não-fumantes, estima-se a expressão do salário hedônico do indivíduo i em função dos seus atributos e de indicadores de risco da atividade j , considerando variáveis instrumentais que possam se correlacionar com a decisão dos indivíduos por atividades de risco e que não sejam relacionados com o termo do erro. Seguindo o mesmo princípio da subseção anterior, o exame do prêmio ao risco estimado para a função condicional média do salário, tal como desenvolvido por [Hersch & Viscusi \(1990\)](#), [Viscusi & Hersch \(2001\)](#), [Esteves \(2008\)](#) e [Contzen, Won & Lavin \(2013\)](#), não medem possíveis heterogeneidades entre os indivíduos, por exemplo, com baixa e alta produtividade, cujos diferenciais compensatórios podem ser diferentes ([FERNANDEZ; NORDMAN, 2009](#); [EVANS; SCHAUR, 2010](#)).

Para levar em conta essa heterogeneidade, as estimações para o prêmio de risco são realizadas, além da tendência central, para funções condicionais quantílicas da produtividade marginal do trabalho, usando a abordagem RQVI. Os quantis 0,10 e 0,25 representam trabalhadores com baixo nível de produtividade marginal do trabalho e os quantis 0,75 e 0,90 os de alta produtividade. Como a variável de risco da indústria (p) é correlacionada com o termo de erro E , a Equação 4.19 expressa os determinantes da demanda dos trabalhadores por atividades arriscadas

$$W_{ij}^s = a_1^s(E_i^s)p_{ij}^s + \mathbf{X}\mathbf{2}_i^s \boldsymbol{\alpha}^s(E_i^s), \text{ com } E_i^s | p_{ij}^s, \mathbf{X}\mathbf{2}_i^s, \mathbf{Z}\mathbf{2}_i^s \sim \text{uniforme}(0, 1) \quad (4.18)$$

$$p_{ij}^s = f(\mathbf{X}\mathbf{2}_i^s, \mathbf{Z}\mathbf{2}_i^s, \Sigma_i^s) \quad (4.19)$$

$$\tau \rightarrow a_1^s(\tau)p_{ij}^s + \mathbf{X}\mathbf{2}_i^s \boldsymbol{\alpha}^s(\tau), \quad (4.20)$$

onde: $s = (f, nf)$; Σ_i^s é um vetor de variáveis omitidas correlacionadas com o termo de erro E_i ; Z é o vetor de variáveis instrumentais baseado em [Garen \(1988\)](#), [Timmins & Murdock \(2007\)](#) e [Contzen, Won & Lavin \(2013\)](#), composto por: TOA (em nível e ao quadrado),

⁴⁹ Como salienta [Fernandez & Nordman \(2009\)](#), além dos benefícios pecuniários, existem outros prêmios não-monetários associados às atividades mais arriscadas.

que representa a quantidade total de ocupados nas outras atividades na mesma divisão econômica, exceto a atividade de atuação do próprio indivíduo; tipo de família e existência de filhos no domicílio; *dummy* se o indivíduo é o chefe do domicílio; rendimento total do não-trabalho; e, cônjuge ou filhos com dificuldades físicas motivados por problemas de saúde. Essas variáveis refletem a atratividade e solvência de um determinado setor e medidas observáveis de aversão ao risco, que segundo [Contzen, Won & Lavin \(2013\)](#), atuam como *proxies* para captar a percepção do trabalhador ao risco de sua atividade e a desejabilidade por um trabalho com mais segurança.

De acordo com a literatura da área, as equações auxiliares no modelo com VI relacionam inicialmente os níveis de perigo de acidentes do trabalho com indicadores observáveis sobre a percepção do nível de risco, que são correlacionados com o risco percebido dos acidentes, mas não explicam o salário recebido pelo indivíduo ([GAREN, 1988](#)). Para [Contzen, Won & Lavin \(2013\)](#), esse bloco inicial de instrumentos consiste em características individuais que captam a aversão ao risco, incluindo renda do não-trabalho e alguns fatores que afetam a percepção de risco dos agentes, funcionando como *proxies* que mostram como os trabalhadores percebem os perigos de suas ocupações. Elas indicam o nível de maturidade do indivíduo quanto ao seu ciclo de vida e, assim, seu desejo por atividades mais seguras.

Por sua vez, o segundo bloco de variáveis exógenas é composto por TOA em nível e ao quadrado, que seriam correlacionados com a medida de risco de cada atividade, mas não correlacionados com a taxa de salário individual. De forma geral, segundo [Timmins & Murdock \(2007\)](#) e [Contzen, Won & Lavin \(2013\)](#), quando trabalhadores escolhem trabalhar em uma determinada firma, eles consideram em algum grau informações sobre as outras firmas. A título de exemplo, [Contzen, Won & Lavin \(2013\)](#) destacam que o tamanho da firma em termos de número de trabalhadores poderia induzir mais pessoas escolherem esta firma, o que impactaria no nível de risco dela (mais trabalhadores afetam o cálculo do índice de risco do setor). Não obstante, o quantitativo de trabalhadores de outras firmas não afetaria, ao menos de imediato, a taxa de salário de indivíduos de uma firma a parte mas que opera no mesmo setor.

Cabe salientar, que os procedimentos de correção pela aproximação da razão inversa de Mills, nos mesmos moldes do retratado na subseção 4.3.1, são empregados para obtenção dos estimadores quantílicos. Com isso, o salário observado dos dois grupos de indivíduos é estabelecido por $S_{ij}^s = d_i^s \times W_{ij}^s$ com a inclusão dos termos expandidos de correção do viés de seleção amostral de [Buchinsky \(1998, 2001\)](#).

4.4 Base de dados e descrição das variáveis

A Pesquisa Nacional por Amostra de Domicílios (PNAD) 2008 é a fonte de informações usada por dispor da Pesquisa Especial de Tabagismo (PETab) e de um suplemento sobre acesso e utilização dos serviços, condições de saúde e fatores de risco e proteção à saúde. Segundo o IBGE (2009), a PETab, desenvolvida em parceria com o Instituto Nacional de Câncer (INCA) e a Agência Nacional de Vigilância Sanitária (ANVISA) vinculados ao Ministério da Saúde, tem o objetivo de levantar dados sobre uso dos produtos derivados do tabaco, das tentativas de abandono do hábito, da percepção das pessoas sobre esses riscos, dos aspectos relacionados à compra de cigarros etc. O questionário da PETab é aplicado a 1/3 da amostra de domicílios da PNAD para 39.425 indivíduos com 15 anos ou mais de idade.

Em termos práticos, o banco de dados e os resultados da pesquisa são divididos em dois estágios, valendo-se de uma amostra específica de observações devido às características dos microdados disponíveis e das necessidades dos modelos econométricos descritos anteriormente, a saber:

- O primeiro sobre a penalização salarial do tabagismo utiliza especificamente a Parte 28 do questionário da PNAD 2008 denominado de PETab, aplicado para uma subamostra de indivíduos do citado *survey*, pela disponibilização de um conjunto mais amplo de variáveis referentes ao vício do cigarro que podem melhor instrumentalizar o modelo empírico. A PETab permite uma melhor análise sobre o uso do tabaco, inclusive considerando o conhecimento dos indivíduos sobre os riscos de saúde provocados pelo uso da nicotina e preços implícitos do produto. O número final de observações é de 6.979 pessoas (quando se realiza o controle amostral para ex-fumantes), o que corresponde em termos populacionais a 24,56 milhões de trabalhadores nas características descritas e a 58% do total da amostra da PETab para esse grupo etário do sexo masculino.
- Na segunda parte, que trata da diferenciação compensatória, uma amostra mais ampla da PNAD é empregada por compatibilizar os questionários da Parte 27, menos detalhado, e da Parte 28, mais minucioso, sobre tabagismo. O maior número de observações é importante para integrar as estatísticas de acidentes do trabalho⁵⁰ do AEAT/MPS com a maior quantidade de atividades econômicas possíveis, onde uma amostra mais restrita resultaria em perdas de dados, reduzindo o conjunto de divisões econômicas avaliadas, bem como o processo de endogeneidade envolvida nesse estágio

⁵⁰ O total de acidentes do trabalho é composto por: acidentes com Comunicações de Acidentes do Trabalho (CAT) registrada; acidentes sem CAT registrada; acidentes típicos, decorrentes das peculiaridades da atividade profissional; acidentes de trajeto; doença do trabalho, produzidas ou desencadeadas pelo exercício do trabalho relativo a determinado ramo de atividade.

é voltada para o índice de risco de acidentes e não para a decisão de fumar. Pela combinação entre a PNAD e o AEAT, o índice de risco de acidente do trabalho⁵¹ (IRAT) é calculado para 54 divisões econômicas (ver Tabela C.1 no Apêndice). As divisões com a maior incidência relativa de acidentes por 100 trabalhadores no período são: Fabricação de máquinas para escritório e equipamentos de informática (9,03); Reciclagem (8,79); e, Fabricação de coque, refino de petróleo, elaboração de combustíveis e nucleares e produção de álcool (4,40).

Com os microdados da PNAD 2008, que possui representatividade amostral para todo o território brasileiro, informações específicas são obtidas sobre os usuários do cigarro, atributos pessoais, aspectos locacionais e indicadores do mercado de trabalho, bem como é possível particionar a amostra em três grupos de trabalhadores: fumantes correntes, ex-fumantes e indivíduos sem experiência com o cigarro. Para Anger & Kvasnicka (2010), essa partição é importante para não contaminar o grupo de não-fumantes com usuários que em algum momento da vida já fizeram uso do tabaco, pelos seguintes pontos: ex-fumantes podem ter um maior salário do que os recém consumidores do cigarro, admitindo que a produtividade é mais correlacionada com as características individuais do que com o uso recente do cigarro; como o tabagismo é viciante, o sucesso em parar de fumar requer alguns atributos que inclusive são relevantes para a produtividade do trabalho, como disciplina e força de vontade.

A exemplo dos outros trabalhos na área, Viscusi & Hersch (2001), Van Ours (2004), Auld (2005), Grafova & Stafford (2009) e Anger & Kvasnicka (2010), a amostra é restringida para trabalhadores do sexo masculino com nacionalidade brasileira e faixa etária entre 25 e 60 anos de idade. Essas restrições têm o propósito de comparar as evidências para o Brasil com os de outros estudos (ver Tabela 4.1), assim como presume-se que aos 25 anos de idade a formação educacional já esteja praticamente maturada ou em fase final e com 60 anos ainda o indivíduo esteja em idade produtiva⁵². Além disso, aposentados e trabalhadores empregados no serviço público foram desconsiderados do desenho amostral. Para o caso dos servidores públicos, a produtividade do trabalho não implica, em regra, alterações na taxa de salário, bem como a existência do fator estabilidade no emprego dificulta o cotejo deste grupo de trabalhadores com os empregados no setor privado⁵³.

⁵¹ Como o índice de risco de acidente de trabalho é dado pela relação entre a quantidade de acidentes e o número de ocupados em cada divisão econômica, as divisões com insuficiência amostral ($n < 30$) ou com falta de informações para compatibilização da Classificação Nacional de Atividades Econômicas (CNAE) domiciliar e com a CNAE 2.0 foram desconsideradas.

⁵² Dada uma expectativa de vida de 69 anos para os homens, conforme dados do Banco Mundial para 2008 disponíveis em databank.worldbank.org/data, bem como o direito a aposentadoria masculina a partir apenas dos 65 anos de idade, de acordo com o art. 48 da lei nº 8.213/1991 (Redação dada pela Lei nº 9.032/1995).

⁵³ Para fins comparativos com os resultados de Viscusi & Hersch (2001), trabalhadores da agricultura e autônomos foram suprimidos da amostra para o modelo de diferenciação compensatória.

Tabela 4.2 – Estatísticas descritivas das variáveis utilizadas nos modelos de penalização e diferenciação compensatória

Variáveis	Penalização				Diferenciação			
	Amostra 1		Amostra 2		Não-fumante		Fumante	
	Média	D.P.	Média	D.P.	Média	D.P.	Média	D.P.
Salário-hora (em log)	2,896	0,896	2,908	0,894	3,060	0,747	2,846	0,704
Sindicalizado	0,215	0,411	0,213	0,410	0,263	0,440	0,201	0,401
<i>Grupos ocupacionais</i>								
Dirigentes	0,070	0,254	0,070	0,255	0,055	0,227	0,031	0,174
Ciências e Artes	0,050	0,217	0,051	0,220	0,079	0,270	0,041	0,199
Técnicos	0,074	0,262	0,075	0,263	0,103	0,304	0,063	0,244
Serviços administrativos	0,051	0,220	0,055	0,228	0,099	0,299	0,067	0,249
Serviços*	0,113	0,317	0,115	0,319	0,153	0,360	0,159	0,366
Vendedores	0,073	0,260	0,072	0,258	0,070	0,255	0,055	0,228
Agrícola	0,168	0,374	0,162	0,368	0,006	0,076	0,014	0,116
Produção e manutenção	0,401	0,490	0,401	0,490	0,436	0,496	0,570	0,495
Estabilidade (<i>tenure</i>)	8,546	9,038	8,123	8,665	5,906	6,452	5,488	6,871
<i>Faixa de escolaridade</i>								
Menos que ensino fundamental (<EF)	0,474	0,499	0,453	0,498	0,269	0,443	0,496	0,500
Ensino fundamental completo (<EM)	0,163	0,369	0,162	0,368	0,181	0,385	0,189	0,392
Ensino médio completo (<ES)	0,275	0,446	0,294	0,456	0,421	0,494	0,253	0,435
Ensino superior completo* (ES)	0,088	0,284	0,092	0,289	0,129	0,335	0,062	0,242
Raça	0,486	0,500	0,490	0,500	0,511	0,500	0,483	0,500
Idade	39,747	9,640	38,589	9,368	35,757	8,339	38,542	9,229
Fumante	0,249	0,432	0,316	0,465	-	-	-	-
Ex-fumante	0,213	0,410	-	-	-	-	-	-
Chefe do domicílio	0,736	0,441	0,721	0,449	0,635	0,481	0,640	0,480
<i>Tipo de família</i>								
Casado sem filho	0,139	0,346	0,136	0,343	0,136	0,343	0,130	0,336
Casado com filho	0,688	0,463	0,682	0,466	0,688	0,463	0,666	0,472
Tamanho da família	3,588	1,494	3,580	1,494	3,507	1,333	3,590	1,466
<i>Rendimento de outras fontes</i>								
Pensão (em log)	0,030	0,427	0,027	0,402	0,020	0,360	0,046	0,529
Aluguel (em log)	0,129	0,887	0,118	0,849	0,086	0,713	0,060	0,605
Outros (em log)	0,137	0,806	0,127	0,777	0,095	0,693	0,100	0,703
Proibido fumar no domicílio	0,137	0,344	0,147	0,354	-	-	-	-
<i>Crença/percepção</i>								
Gera doenças aos outros	0,920	0,272	0,913	0,281	-	-	-	-
Gera doenças para si	0,966	0,182	0,962	0,191	-	-	-	-
Fumantes na família	0,123	0,404	0,149	0,442	-	-	-	-
Ex-fumantes na família	0,074	0,261	0,020	0,140	-	-	-	-
Preço do cigarro	2,648	0,476	2,646	0,473	-	-	-	-
Renda total não-trabalho (em log)	-	-	-	-	0,198	1,039	0,200	1,044
TOA	-	-	-	-	2,797	4,018	2,439	3,838
Urbana	0,836	0,370	0,840	0,367	0,946	0,226	0,934	0,249
Região metropolitana	0,400	0,490	0,398	0,489	0,380	0,485	0,389	0,488
<i>Macrorregião</i>								
Norte	0,079	0,270	0,076	0,265	0,063	0,243	0,070	0,256
Nordeste	0,242	0,428	0,244	0,429	0,211	0,408	0,178	0,383
Sudeste*	0,450	0,498	0,453	0,498	0,514	0,500	0,506	0,500
Sul	0,157	0,364	0,157	0,363	0,142	0,349	0,176	0,381
Centro-Oeste	0,072	0,259	0,071	0,257	0,070	0,255	0,069	0,254
Controle: Ex-fumantes			X		X		X	
Observações (N)	8.808		6.979		22.088		8.153	
Observações expandidas** (P)	31.221.094		24.561.120		10.739.467		4.080.790	

Fonte: Elaboração própria a partir dos microdados da PNAD/PETab 2008.

*Categoria omitida.

**As amostras possuem fatores de expansão distintos, dada às peculiaridades da PETab que é uma subamostra da PNAD com reponderação do peso amostral dos participantes.

Enfatiza-se que as seguintes variáveis explicativas têm interseções nos dois modelos: faixa de escolaridade (menos que ensino fundamental, fundamental completo, médio completo e superior completo), raça (cor branca), idade (em nível e ao quadrado), interação entre anos de estudo e idade do indivíduo, *dummy* se o trabalhador é sindicalizado, estabilidade no emprego atual (*tenure*) e fatores locais (zona de localização, região

metropolitana e macrorregiões de residência). No primeiro modelo é acrescentado variáveis sobre os grupos ocupacionais do trabalho principal⁵⁴ e a variável de interesse sobre o status de fumante, ao tempo que o segundo modelo segmenta a equação de rendimentos e inclui um índice de risco (IRAT) como *proxy* para a probabilidade de ocorrência de danos à saúde do trabalhador em decorrência do exercício de seu trabalho. A Tabela C.3 no Apêndice mostra a definição das variáveis utilizadas no estudo, inclusive as variáveis instrumentais e em qual modelagem as mesmas são utilizadas. Por sua vez, a Tabela 4.2 sumariza as estatísticas descritivas das variáveis usadas para os modelos de penalização para dois tipos de amostras, que se diferenciam pelo controle de ex-fumantes, e para o de diferenciação compensatória com dados para fumantes e não-fumantes. No modelo de diferenciação compensatória a amostra de não-fumantes desconsidera ex-usuários do cigarro.

Pela Tabela 4.2, a amostra extraída da PETab, para o modelo de penalização, dispõe de dois desenhos amostrais que se diferenciam pelo controle de ex-fumantes. Na Amostra (1), 21,3% dos indivíduos são ex-usuários do cigarro, enquanto que na Amostra (2) essas observações são excluídas. A comparação dessas duas amostras não revela grandes diferenças. Ambas são caracterizadas por pessoas residentes, em sua maior parte, na região sudeste, em áreas urbanas, com ensino fundamental incompleto, casadas e com idade em torno de 39 anos. Em relação à amostra para o modelo de diferenciação compensatória, a média de idade dos trabalhadores não-fumantes é de 36 anos e dos fumantes de 39 anos, com os primeiros apresentando maiores níveis do log dos salários e com uma menor concentração de indivíduos com menor faixa de escolaridade. A maior parte dos fumantes, com base no perfil de indivíduos (homens entre 25 e 60 anos) e outros recortes amostrais, está localizada nas regiões Sudeste (50,6%), Nordeste (17,8%) e Sul (17,6%).

4.5 Resultados

Os resultados estão decompostos em duas partes centrais: penalização do salário motivada pelo uso do cigarro e diferenciação compensatória salarial para atividades perigosas entre fumantes e não-fumantes.

4.5.1 Penalização salarial para usuários do cigarro

Conforme Anger & Kvasnicka (2010), a verificação da hipótese de penalização salarial do tabagismo requer cuidados específicos com o grupo de não-fumantes, porque os custos salariais efetivos de fumar podem ser influenciados pela existência de ex-usuários do cigarro neste grupo. Assim, para iniciar as discussões, a Tabela 4.3 mostra as características

⁵⁴ No caso do modelo de diferenciação compensatória, a inclusão da variável IRAT inclui ao modelo os diferentes níveis de risco de acidentes do trabalho por divisão econômica de ocupação dos trabalhadores, por isso a não inclusão dos grupos ocupacionais.

da amostra por grupos de fumantes e não-fumantes de dois tipos, diferenciando-se pelo o uso passado do cigarro. Dentre as diferenças de médias e proporções entre os grupos, apenas quatro não são estatisticamente diferentes de zero à pelo menos 5% de significância: renda do não-trabalho e trabalhador casado (com filho) entre fumantes (A) e não-fumantes do tipo 1 (B); horas de trabalho e morador chefe do domicílio entre fumantes (A) e ex-fumantes ou não-fumantes do tipo 2 (C). Na Tabela 4.3 as diferenças de médias e de proporções são sempre examinadas em torno do grupo de fumantes.

Tabela 4.3 – Características da amostra por grupos de fumantes e não-fumantes no Brasil, PETab 2008

Variáveis	Fumante (A)	Não-Fumante		Diferença			
		Tipo 1 (B)	Tipo 2 (C)	Coef. (A)-(B)	Estat. t (A)-(B)	Coef. (A)-(C)	Estat. t (A)-(C)
Média							
Salário-hora	24,39	33,14	27,57	-8,74	4,10	-3,18	3,50
Horas de trabalho	44,45	44,55	45,13	-0,10	6,49	-0,67	1,52
<i>Tenure</i> (em anos)	8,15	7,26	9,18	0,89	15,78	-1,03	4,22
Anos de estudo	5,77	8,19	6,57	-2,41	34,87	-0,79	5,31
Idade	41,44	37,38	44,16	4,06	26,25	-2,71	19,20
Renda do não-trabalho	29,48	26,28	40,61	3,20	1,68	-11,13	3,91
Idade de 1º uso do cigarro	16,37		15,89			0,48	2,61
Número de cigarros ao dia	17,74						
Proporção							
Ocupação	88,89%	91,12%	91,05%	-2,23%	12,38	-2,16%	4,12
Chefe do domicílio	72,67%	66,75%	77,30%	5,92%	26,90	-4,63%	1,34
Cor Branca	42,23%	51,10%	46,43%	-8,87%	7,60	-4,20%	4,02
Ausência no trabalho*	9,32%	5,09%	9,62%	4,23%	5,57	-0,30%	3,95
Dificuldades físicas	21,28%	12,75%	23,28%	8,53%	12,58	-2,00%	8,28
Casado e sem filhos	14,51%	12,61%	15,09%	1,90%	1,70	-0,58%	5,10
Casado e com filhos	63,92%	68,72%	69,35%	-4,80%	11,37	-5,43%	3,16
<i>Faixa de escolaridade</i>							
<EF	63,54%	38,10%	55,46%	25,44%	31,15	8,08%	4,04
<EM	15,04%	16,87%	16,31%	-1,83%	6,74	-1,27%	2,17
<ES	16,73%	33,72%	20,35%	-16,99%	21,05	-3,62%	2,89
ES	4,70%	11,31%	7,88%	-6,61%	10,30	-3,18%	6,84

Fonte: Elaboração própria a partir dos microdados da PNAD/PETab 2008. Amostra restringida para homens nascidos no Brasil com faixa etária entre 25 e 60 anos, com exclusão de indivíduos que recebem aposentadoria e trabalhadores ocupados no serviço público.

Não-fumante Tipo 1: indivíduo que nunca teve experiência com o uso do cigarro.

Não-fumante Tipo 2: indivíduo ex-fumante.

*Essa variável indica se nas duas últimas semanas, a pessoa não fez suas atividades habituais por motivo de saúde.

Pela Tabela 4.3, fumantes e não-fumantes dispõem de certas características socioeconômicas distintas, principalmente nas variáveis anos de estudo⁵⁵ (e na sua distribuição por fase de ensino) e ganhos no mercado de trabalho. Essas diferenças são mais acentuadas quando se coteja a média desses indicadores entre fumantes e não-fumantes sem experiência com o cigarro. Enquanto que variáveis como horas de trabalho, estabilidade no emprego (*tenure*), renda do não-trabalho (exceto para os ex-fumantes), taxa de ocupação e outros atributos familiares relativamente não exibem diferenças tão marcantes como as duas supracitadas variáveis (educação e salário). Observando o salário-hora, na média, os fumantes recebem menos que os dois tipos de não-fumantes, com diferenças estatisticamente

⁵⁵ Existe uma linha de trabalhos que investigam as diferentes formas de causalidade envolvendo educação e saúde, visto que o hábito de fumar pode revelar altas taxas de descontos intertemporais, que por sua vez pode resultar em baixos investimentos na qualificação. Maiores detalhes sobre esse tipo de discussão, por exemplo, em [Conti, Heckman & Urzua \(2010\)](#) e [Cutler & Lleras-Muney \(2010\)](#)

significativas a pelo menos 1%. Sem condicionar a outros controles, o diferencial salarial entre fumantes e não-usuários do cigarro do tipo 1 é de R\$ 8,74 por hora de trabalho em favor dos não-fumantes.

Em conformidade com as características da amostra, a média de idade que os fumantes e ex-fumantes consumiram o cigarro pela primeira vez é de cerca de 16 anos, bem como o número médio de cigarros consumidos diariamente pelos atuais fumantes é de quase um maço de cigarro. Logo, uma iniciação prematura ao fumo combinada com uma elevada quantidade consumida dele ao dia podem ter um impacto expressivo no estoque de saúde dos fumantes, que com apenas 25 anos de idade, com base nesses prognósticos para a média, já teriam consumidos pelo menos 53,97 mil cigarros. Assim, a Tabela 4.3 reporta que fumantes e ex-fumantes possuem uma maior taxa de ausência no trabalho por motivos de saúde, quase o dobro da registrada para os não-fumantes que nunca consumiram o cigarro. Esse fato pode traduzir em uma maior tendência de absenteísmo do local do trabalho para esse grupo de pessoas, refletindo em um menor retorno salarial (LEVINE; GUSTAFSON; VELENCHIK, 1997).

A existência de dificuldade física para levantar objetos, praticar esportes ou realizar trabalhos pesados é outro ponto de disparidade, dada uma taxa de 8,5 pontos percentuais (p.p.) maior para os fumantes no comparativo com os não-fumantes tipo 1, tendo os ex-fumantes a maior proporção de indivíduos com essa dificuldade (2% a mais que os fumantes). Considerando uma possível contaminação no grupo de não-fumantes, que contém indivíduos com experiência pretérita de uso do cigarro – que, em média, aproximam-se mais dos fumantes em características observadas na Tabela 4.3 –, os modelos econométricos fazem filtro amostral para ex-fumantes⁵⁶.

Na finalidade de adicionar a heterogeneidade da penalização salarial provocada pelo uso do cigarro sem perder de vista os problemas de endogeneidade, na parte 4.5.1.2 são apresentados os resultados da pesquisa por meio do método RQVI. Contudo, a subseção 4.5.1.1, como ponto de partida, expõe as estimativas calculadas para a função condicional média dos salários.

4.5.1.1 Resultados por MQO e MQ2E

Os coeficientes do modelo de regressão para a penalização salarial do tabagismo estão exibidos na Tabela 4.4. Essas estimativas servem como referência para os resultados obtidos através dos modelos com estimadores quantílicos. Do modelo (1) ao (4) são expostos os coeficientes por MQO sem e com controle na amostra para ex-fumantes. As duas últimas colunas da tabela exibem as estimativas por MQ2E, diferenciando-se apenas no controle amostral para ex-fumantes.

⁵⁶ Esse filtro não induz a mudanças drásticas nos achados de penalização, como pode ser visualizado na Tabela 4.4, mas é utilizado para dar um maior rigor na análise.

Tabela 4.4 – Resultados da regressão para a penalização salarial do fumante corrente no Brasil, PETab 2008. Variável dependente: logaritmo do salário-hora

Covariadas	MQO				MQ2E	
	(1)	(2)	(3)	(4)	(5)	(6)
Fumante	-0,3450*** (0,0272)	-0,1258*** (0,0240)	-0,1053*** (0,0212)	-0,0885*** (0,0195)	-0,2160*** (0,0463)	-0,2479*** (0,0526)
Faixa de escolaridade						
<EF		-0,6193*** (0,0759)	-0,3778*** (0,0843)	-0,3751*** (0,0728)	-0,3658*** (0,0730)	-0,3648*** (0,0844)
<EM		-0,6248*** (0,0582)	-0,3818*** (0,0682)	-0,3730*** (0,0593)	-0,3704*** (0,0596)	-0,3772*** (0,0686)
<ES		-0,5223*** (0,0494)	-0,3300*** (0,0584)	-0,3180*** (0,0515)	-0,3203*** (0,0520)	-0,3334*** (0,0591)
Raça		0,2249*** (0,0213)	0,1020*** (0,0201)	0,1043*** (0,0179)	0,1014*** (0,0179)	0,0971*** (0,0201)
Idade		0,0409*** (0,0102)	0,0308*** (0,0096)	0,0281*** (0,0084)	0,0286*** (0,0084)	0,0306*** (0,0095)
Idade2/100		-0,0515*** (0,0124)	-0,0401*** (0,0116)	-0,0375*** (0,0100)	-0,0376*** (0,0100)	-0,0382*** (0,0116)
Interação: Educação e Idade		0,0020*** (0,0001)	0,0011*** (0,0001)	0,0011*** (0,0001)	0,0011*** (0,0001)	0,0011*** (0,0001)
Sindicalizado			0,1248*** (0,0238)	0,1198*** (0,0208)	0,1130*** (0,0208)	0,1156*** (0,0238)
Estabilidade (tenure)			0,0095*** (0,0014)	0,0099*** (0,0012)	0,0098*** (0,0012)	0,0094*** (0,0014)
Grupos ocupacionais						
Dirigentes			0,7921*** (0,0544)	0,7853*** (0,0470)	0,7813*** (0,0477)	0,7881*** (0,0553)
Ciências e Artes			0,7320*** (0,0809)	0,7311*** (0,0691)	0,7221*** (0,0699)	0,7221*** (0,0819)
Técnicos			0,5782*** (0,0440)	0,5788*** (0,0412)	0,5774*** (0,0413)	0,5789*** (0,0442)
Serviços administrativos			0,1672*** (0,0408)	0,1916*** (0,0366)	0,1869*** (0,0369)	0,1618*** (0,0411)
Vendedores			0,1679*** (0,0379)	0,1381*** (0,0377)	0,1313*** (0,0380)	0,1603*** (0,0384)
Agrícola			-0,2135*** (0,0396)	-0,2116*** (0,0356)	-0,2061*** (0,0358)	-0,2058*** (0,0398)
Produção e manutenção			0,1582*** (0,0243)	0,1797*** (0,0216)	0,1817*** (0,0217)	0,1616*** (0,0246)
Fatores locacionais						
Urbana			0,0985** (0,0389)	0,1160*** (0,0340)	0,1175*** (0,0339)	0,1008*** (0,0387)
Região metropolitana			-0,0236 (0,0187)	-0,0244 (0,0169)	-0,0242 (0,0169)	-0,0228 (0,0187)
Norte			-0,1569*** (0,0323)	-0,1272*** (0,0281)	-0,1277*** (0,0281)	-0,1550*** (0,0325)
Nordeste			-0,4062*** (0,0254)	-0,3921*** (0,0225)	-0,3943*** (0,0225)	-0,4096*** (0,0254)
Sul			-0,0053 (0,0273)	0,0104 (0,0243)	0,0114 (0,0243)	-0,0035 (0,0272)
Centro-Oeste			-0,0085 (0,0288)	0,0274 (0,0273)	0,0241 (0,0274)	-0,0118 (0,0291)
λ			-0,4258*** (0,0822)	-0,4383*** (0,0748)	-0,4251*** (0,0749)	-0,4041*** (0,0822)
Constante	3,0164*** (0,0153)	2,0441*** (0,1957)	2,1590*** (0,1840)	2,1750*** (0,1648)	2,1867*** (0,1653)	2,1830*** (0,1847)
Controle: ex-fumantes	X	X	X		X	X
Instrumentos						
N	6,979	6,979	6,979	8,808	8,808	6,979
P (em milhões)	24,56	24,56	24,56	31,22	31,22	24,56
R ² ajustado	0,03	0,37	0,49	0,47	0,47	0,49

Fonte: Elaboração própria a partir dos microdados da PNAD/PETab 2008. Erros-padrão robustos entre parênteses, corrigidos por correlação intra-grupo e independência inter-grupos.

Instrumentos: proibido fumar no domicílio, crença/percepção sobre as consequências do cigarro (gera doenças aos outros e para si), fumantes na família, existência de ex-fumantes na família e preço do cigarro.

Nota: O λ nesses modelos representa a razão inversa de Mills tradicional proposta por Heckman (1979).

Legenda: *p<0,10; **p<0,05; ***p<0,01.

Em termos de ajustamento dos modelos, as variáveis explicativas comumente utilizadas pela literatura da área, escolaridade, idade (que capta o efeito da experiência) e outros atributos individuais e locacionais, mostram-se com os sinais esperados e com significância estatística. Trabalhadores com menos escolaridade e situados nas regiões Norte e Nordeste apresentam uma menor remuneração, enquanto que indivíduos de cor branca, sindicalizados, com maior tempo de serviço na mesma empresa, residentes em áreas urbanas e que ocupam, em especial, cargos de dirigentes e atuam como profissionais nas áreas das ciências e das artes possuem maiores taxas de salários. Além do mais, o grau de determinação do modelo, R^2 ajustado, é de pelo menos 47%, bem como a razão inversa de Mills tradicional (λ) tem significância estatística nas especificações em que ela é incluída, sinalizando a importância da correção do viés de seleção. O $\lambda < 0$ expressa que os fatores latentes na equação de ganhos minceriana, que se relacionam com uma maior probabilidade de participação na força de trabalho, diminuem a taxa de retornos salariais.

O parâmetro associado aos indivíduos fumantes nos modelos sem VI é negativo e significativo estatisticamente para todas as especificações. Ademais, no modelo (1), que não controla os atributos pessoais, ocupacionais e locacionais dos indivíduos, o coeficiente do fumante é de -0,345, que é atenuado em termos absolutos ao ponderar os demais conjuntos de variáveis explicativas, situando-se em -0,105 no modelo (3). Muito embora, o acréscimo das características individuais, ocupacionais e locacionais reduzirem os efeitos do tabagismo no logaritmo do salário-hora, os custos do tabagismo sobre os rendimentos individuais podem ser tendenciosos caso variáveis não-observadas se correlacionem com os salários e com a decisão de fumar.

Após o controle de endogeneidade, nota-se que o fumante ganha em média 19,4% a menos que um não-fumante⁵⁷, uma taxa de desconto bem acima das estimativas desenvolvidas por MQO (respectivamente, 10% e 8,5% nos modelos 3 e 4), indicando a existência em valores absolutos de uma possível subestimação no citado efeito pela regressão clássica, inclusive do erro-padrão das estimativas. Tais achados seguem a tendência da literatura internacional, como ilustrado na Tabela 4.1. Em Auld (2005), por exemplo, após o controle para endogeneidade, o uso do cigarro provoca um impacto de -24% no salário de trabalhadores canadenses com baixa sensibilidade a escolha dos instrumentos. Na presente pesquisa, as diferentes especificações de VI, apresentadas no Painel A da Tabela C.4 no Apêndice, mostram que o intervalo para a penalização para o caso nacional situa-se entre -0,30 e -0,22.

A amostra contaminada pelos ex-fumantes reduz também os efeitos da penalização salarial. Ao excluir da base amostral os ex-fumantes, ocorre uma ampliação da penalização

⁵⁷ Baseado em Halvorsen & Palmquist (1980), o valor da semielasticidade para o coeficiente do fumante (por exemplo, -0,216), que é um regressor binário, é calculado daqui em diante por: $(e^{-0,216} - 1) \times 100 = 19,4\%$.

salarial em 1,52 p.p. entre os modelos (3) e (4) e de 2,53 p.p. entre os modelos (6) e (5). Como os ex-fumantes recebem, em média, uma remuneração inferior do que aqueles que nunca fumaram, a base de comparação tende a subestimar os custos salariais do hábito de fumar. Desta maneira, a permanência dos ex-fumantes e o não controle para endogeneidade geram uma tendência para baixo no valor absoluto do coeficiente da penalização, distorcendo os custos monetários implícitos do uso do cigarro⁵⁸.

Quanto ao MQ2E, como pode ser verificado na Tabela 4.5, os resultados condicionados ao vetor de instrumentos exibem um ajuste aceitável na determinação da decisão de fumar do indivíduo, de acordo com Stock, Wright & Yogo (2002), Stock & Yogo (2005). No tratamento do problema de endogeneidade, três testes são comumente usados para mensurar o ajustamento da modelagem com VI: (i) teste de endogeneidade; (ii) condições de restrições de sobre-identificação dos instrumentos; e, (iii) resistência dos instrumentos. A Tabela 4.5 reporta as estatísticas dos supracitados testes.

Inicialmente, embasado na amostra em estudo, a decisão de fumar é endógena? Caso ela seja exógena, o resultado por MQO seria mais eficiente do que a abordagem de VI (WOOLDRIDGE, 2010). Pelo teste de endogeneidade, a hipótese de exogeneidade para a variável H_i não pode ser aceita, em ambos os testes (Durbin e Wu-Hausman). Com isso, os resultados dos modelos (1) ao (4), supondo a sua exogeneidade, são tendenciosos. Outro teste importante é o de análise de sobre-identificação das restrições, para situações em que a quantidade de instrumentos supera o número de covariadas endógenas. Esse teste admite a validade de um instrumento para em seguida observar a validade dos demais (ou seja, se os instrumentos não estão correlacionados com o termo de erro na segunda etapa). Sendo assim, a hipótese de sobre-identificação é rejeitada pelos valores estatísticos dos testes de restrição.

Para uma variável ser um instrumento válido, ela deve ser suficientemente correlacionada com o regressor endógeno mas não correlacionada com o termo de erro. Na Tabela 4.5 estão disponíveis estatísticas sobre o poder explicativo dos instrumentos. Pelo ajustamento do primeiro estágio através do teste F, todos os instrumentos apresentam, de forma conjunta, significância estatística (visto o p-valor = 0), bem como o valor de $F > 11$ indica que os instrumentos são suficientemente confiáveis a pelo menos um viés relativo de 10% (STOCK; WRIGHT; YOGO, 2002). Conforme os valores críticos calculados a partir do teste de identificação de Stock & Yogo (2005), as estatísticas de Cragg-Donald e Kleibergen-Paap estão bem acima dos pontos críticos, implicando que os instrumentos não podem ser taxados como fracos. Ademais, os valores para o R^2 constantes na Tabela

⁵⁸ A Tabela C.2 no Apêndice reporta outra especificação do modelo de penalização salarial, incorporando covariadas sobre os ramos da atividade e ocupação informal para os modelos 3 e 6 da Tabela 4.4. Com base nesses resultados, o coeficiente para os fumantes permanecem negativos e com uma variação de -0,105 no modelo 3 para -0,094 no modelo 3' e de -0,248 no modelo 6 para -0,23 no modelo 6', o que de certa maneira mostra a robustez dos resultados frente a diferentes especificações.

4.5 apresentam valores maiores que 20%, ratificando o grau de ajuste dos instrumentos utilizados para identificação dos fumantes.

Tabela 4.5 – Testes para endogeneidade da variável H_i no MQ2E

Teste	Modelo (5)		Modelo (6)	
	Estatística	P-valor	Estatística	P-valor
Endogeneidade ($H_0: H_i$ exógeno)				
Escore robusto (Durbin) $\chi^2_{(1)}$	18,742	0,000	18,148	0,000
Regressão robusta (Wu-Hausman) F	10,464	0,001	9,927	0,002
Restrição de sobre-identificação				
Escore $\chi^2_{(5)}$	274,940	0,000	192,732	0,000
Estatística J de Hansen: $\chi^2_{(5)}$	232,570	0,000	165,203	0,000
Resistência dos instrumentos				
<i>Ajustamento do primeiro estágio</i>				
R^2	0,2663		0,287	
R^2 ajustado	0,264		0,284	
R^2 parcial	0,231		0,209	
F robusto	182,894	0,000	111,081	0,000
R^2 parcial de Shea	0,2308		0,2088	
R^2 parcial ajustado de Shea	0,228		0,205	
<i>Teste de identificação de instrumentos fracos</i>				
Cragg-Donald (Estatística F de Wald)	439,076		305,635	
Kleibergen-Paap (Estatística F de Wald)	182,894		111,081	
Valores críticos do teste de Stock & Yogo (2005)				
Viés relativo máximo de 5% na VI	19,28		19,28	
Viés relativo máximo de 10% na VI	11,12		11,12	
Viés relativo máximo de 20% na VI	6,76		6,76	
Viés relativo máximo de 30% na VI	5,15		5,15	
Tamanho máximo de 10% na VI	29,18		29,18	
Tamanho máximo de 15% na VI	16,23		16,23	
Tamanho máximo de 20% na VI	11,72		11,72	
Tamanho máximo de 25% na VI	9,38		9,38	

Fonte: Elaboração própria a partir dos microdados da PNAD/PETab 2008.

4.5.1.2 Resultados por Regressão quantílica e RQVI

Os efeitos do cigarro sobre o salário por MQO e MQ2E revelam apenas a tendência central da relação entre essas duas variáveis, não captando possíveis heterogeneidades do efeito ao longo da distribuição condicional dos salários. Os coeficientes da penalização salarial do fumante corrente através de estimadores quantílicos com controle da amostra para ex-fumantes estão reportados nas Tabelas 4.6 e 4.7, com utilização da abordagem RQVI na última tabela. Os valores de τ estão dispostos para os quantis: 0,10; 0,25; 0,50; 0,75 e 0,90.

Tabela 4.6 – Resultados para a penalização salarial do fumante corrente por RQ no Brasil, PETab 2008. Variável dependente: logaritmo do salário-hora

Covariadas	Q(0,10)	Q(0,25)	Q(0,50)	Q(0,75)	Q(0,90)
Fumante	-0,1296*** (0,0263)	-0,1013*** (0,0189)	-0,0743*** (0,0166)	-0,0678*** (0,0176)	-0,1006*** (0,0238)
Faixa de escolaridade					
<EF	-0,2462*** (0,0849)	-0,3709*** (0,0716)	-0,5000*** (0,0729)	-0,5746*** (0,0702)	-0,3895*** (0,0949)
<EM	-0,2551*** (0,0567)	-0,3765*** (0,0609)	-0,4980*** (0,0641)	-0,5972*** (0,0590)	-0,4081*** (0,0777)
<ES	-0,2365*** (0,0411)	-0,3103*** (0,0544)	-0,4137*** (0,0599)	-0,4802*** (0,0541)	-0,2984*** (0,0683)
Raça	0,0335 (0,0246)	0,0414** (0,0163)	0,0617*** (0,0160)	0,0678*** (0,0176)	0,1505*** (0,0296)
Idade	0,0197* (0,0103)	0,0244*** (0,0084)	0,0266*** (0,0075)	0,0221** (0,0086)	0,0283** (0,0123)
Idade ² /100	-0,0385*** (0,0122)	-0,0361*** (0,0106)	-0,0353*** (0,0092)	-0,0285*** (0,0108)	-0,0352** (0,0158)
Interação: Educação e Idade	0,0012*** (0,0002)	0,0009*** (0,0001)	0,0010*** (0,0001)	0,0011*** (0,0001)	0,0014*** (0,0002)
Sindicalizado	0,1345*** (0,0256)	0,1355*** (0,0196)	0,1301*** (0,0165)	0,1353*** (0,0231)	0,1473*** (0,0343)
Estabilidade (<i>tenure</i>)	0,0086*** (0,0015)	0,0066*** (0,0012)	0,0081*** (0,0011)	0,0115*** (0,0013)	0,0161*** (0,0018)
Grupos ocupacionais					
Dirigentes	0,3199*** (0,0744)	0,6331*** (0,0560)	0,7927*** (0,0407)	0,9953*** (0,0527)	1,1982*** (0,0758)
Ciências e Artes	0,4939*** (0,0741)	0,6645*** (0,0616)	0,7456*** (0,0544)	0,8331*** (0,0735)	1,0858*** (0,1139)
Técnicos	0,3021*** (0,0389)	0,4379*** (0,0416)	0,5636*** (0,0462)	0,6625*** (0,0444)	0,8549*** (0,0814)
Serviços administrativos	0,0675* (0,0410)	0,1658*** (0,0388)	0,2056*** (0,0311)	0,1798*** (0,0388)	0,1642*** (0,0395)
Vendedores	0,0325 (0,0428)	0,0586** (0,0241)	0,0840*** (0,0316)	0,2164*** (0,0607)	0,3329*** (0,0660)
Agrícola	-0,5108*** (0,0645)	-0,2848*** (0,0377)	-0,1832*** (0,0312)	-0,0550* (0,0331)	0,0073 (0,0523)
Produção e manutenção	0,0349 (0,0358)	0,1225*** (0,0234)	0,1521*** (0,0203)	0,1699*** (0,0204)	0,1685*** (0,0349)
Fatores locacionais					
Urbana	0,1765*** (0,0461)	0,1092*** (0,0221)	0,0796*** (0,0255)	0,1102*** (0,0307)	0,1120** (0,0455)
Região metropolitana	-0,0064 (0,0229)	-0,0033 (0,0157)	-0,0184 (0,0154)	-0,0443*** (0,0165)	-0,0577** (0,0247)
Norte	-0,2059*** (0,0411)	-0,2158*** (0,0260)	-0,1697*** (0,0263)	-0,1281*** (0,0267)	-0,0873* (0,0517)
Nordeste	-0,4869*** (0,0345)	-0,4162*** (0,0210)	-0,3473*** (0,0191)	-0,3286*** (0,0226)	-0,3084*** (0,0275)
Sul	0,0107 (0,0256)	0,0190 (0,0242)	0,0185 (0,0236)	0,0245 (0,0259)	0,0124 (0,0345)
Centro-Oeste	-0,0332 (0,0383)	-0,0605** (0,0241)	0,0154 (0,0256)	0,0258 (0,0252)	-0,0004 (0,0514)
h_1	-0,0829 (0,1371)	-0,0772** (0,0385)	-0,0995 (0,0689)	-0,0653** (0,0284)	-0,0445 (0,0415)
h_2	-0,4396** (0,2001)	-0,3783*** (0,0885)	-0,2678** (0,1089)	-0,3030*** (0,0483)	-0,1137 (0,0836)
Constante	1,8803*** (0,2038)	2,1128*** (0,1663)	2,3600*** (0,1565)	2,6979*** (0,1721)	2,5489*** (0,2405)
Controle amostral: ex-fumantes	X	X	X	X	X
Instrumentos					
N	6.979	6.979	6.979	6.979	6.979
P (em milhões)	24,56	24,56	24,56	24,56	24,56
Pseudo-R ²	0,2634	0,2475	0,3033	0,3588	0,3718

Fonte: Elaboração própria a partir dos microdados da PNAD/PETab 2008. Erros-padrão robustos entre parênteses, corrigidos por correlação intra-grupo e independência inter-grupos.

Nota: h_1 e h_2 representam os dois termos de correção do viés de seleção para estimadores quantílicos.

Legenda: *p<0,10; **p<0,05; ***p<0,01

Tabela 4.7 – Resultados para a penalização salarial do fumante corrente por RQVI no Brasil, PETab 2008. Variável dependente: logaritmo do salário-hora

Covariadas	Q(0,10)	Q(0,25)	Q(0,50)	Q(0,75)	Q(0,90)
Fumante	-0,1644*** (0,0597)	-0,2310*** (0,0371)	-0,2019*** (0,0358)	-0,1766*** (0,0391)	-0,4547*** (0,0664)
Faixa de escolaridade					
<EF	-0,2866*** (0,0924)	-0,3376*** (0,0746)	-0,4669*** (0,0688)	-0,5014*** (0,0716)	-0,3641*** (0,1098)
<EM	-0,2865*** (0,0699)	-0,3412*** (0,0645)	-0,4482*** (0,0584)	-0,5155*** (0,0618)	-0,4163*** (0,0975)
<ES	-0,2383*** (0,0582)	-0,2867*** (0,0588)	-0,3847*** (0,0518)	-0,4364*** (0,0543)	-0,3476*** (0,0840)
Raça	0,0358 (0,0252)	0,0375** (0,0173)	0,0785*** (0,0173)	0,0640*** (0,0195)	0,1525*** (0,0301)
Idade	0,0232* (0,0120)	0,0208** (0,0090)	0,0281*** (0,0074)	0,0262*** (0,0093)	0,0358*** (0,0126)
Idade ² /100	-0,0403*** (0,0149)	-0,0298*** (0,0113)	-0,0352*** (0,0092)	-0,0333*** (0,0114)	-0,0412*** (0,0155)
Interação: Educação e Idade	0,0010*** (0,0002)	0,0008*** (0,0001)	0,0008*** (0,0001)	0,0010*** (0,0001)	0,0012*** (0,0002)
Sindicalizado	0,1163*** (0,0300)	0,1294*** (0,0207)	0,1132*** (0,0190)	0,1292*** (0,0252)	0,1107*** (0,0357)
Estabilidade (<i>tenure</i>)	0,0085*** (0,0018)	0,0064*** (0,0013)	0,0081*** (0,0012)	0,0108*** (0,0014)	0,0114*** (0,0022)
Grupos ocupacionais					
Dirigentes	0,3274*** (0,0625)	0,6452*** (0,0557)	0,7639*** (0,0410)	0,9185*** (0,0548)	1,1071*** (0,0963)
Ciências e Artes	0,4925*** (0,0746)	0,6669*** (0,0664)	0,7427*** (0,0614)	0,8197*** (0,0775)	0,9251*** (0,1101)
Técnicos	0,3268*** (0,0482)	0,4372*** (0,0381)	0,5396*** (0,0446)	0,6334*** (0,0469)	0,7719*** (0,0738)
Serviços administrativos	0,0995** (0,0459)	0,1554*** (0,0440)	0,2160*** (0,0357)	0,1853*** (0,0428)	0,1010* (0,0597)
Vendedores	0,0375 (0,0458)	0,0372 (0,0298)	0,0740** (0,0347)	0,2053*** (0,0511)	0,3633*** (0,0719)
Agrícola	-0,4976*** (0,0725)	-0,2867*** (0,0429)	-0,1611*** (0,0343)	-0,0631* (0,0356)	0,0474 (0,0513)
Produção e manutenção	0,0541 (0,0343)	0,1191*** (0,0225)	0,1582*** (0,0214)	0,1809*** (0,0244)	0,1766*** (0,0399)
Fatores locacionais					
Urbana	0,1519*** (0,0521)	0,0786** (0,0319)	0,0934*** (0,0288)	0,1097*** (0,0290)	0,1204*** (0,0391)
Região metropolitana	-0,0163 (0,0250)	-0,0079 (0,0170)	-0,0333** (0,0159)	-0,0497*** (0,0187)	-0,0308 (0,0303)
Norte	-0,2432*** (0,0404)	-0,2173*** (0,0297)	-0,1490*** (0,0271)	-0,1464*** (0,0319)	-0,0795* (0,0452)
Nordeste	-0,5032*** (0,0385)	-0,4046*** (0,0222)	-0,3493*** (0,0202)	-0,3211*** (0,0244)	-0,2837*** (0,0353)
Sul	0,0391 (0,0348)	0,0432* (0,0242)	0,0740*** (0,0248)	0,0832*** (0,0271)	0,1228*** (0,0459)
Centro-Oeste	-0,0653* (0,0371)	-0,0407 (0,0271)	0,0224 (0,0280)	0,0390 (0,0302)	0,0505 (0,0509)
h_1	-0,0091 (0,0426)	-0,0665* (0,0356)	-0,0540 (0,0528)	-0,0473 (0,0910)	-0,0618 (0,1728)
h_2	-0,5403*** (0,0645)	-0,4032*** (0,0614)	-0,3081*** (0,0889)	-0,3448** (0,1543)	-0,2669 (0,3062)
Constante	1,4591*** (0,2634)	1,7762*** (0,1967)	1,8458*** (0,1717)	1,9957*** (0,2074)	1,4620*** (0,2996)
Controle amostral: ex-fumantes	X	X	X	X	X
Instrumentos	X	X	X	X	X
N	6.979	6.979	6.979	6.979	6.979
P (em milhões)	24,56	24,56	24,56	24,56	24,56

Fonte: Elaboração própria a partir dos microdados da PNAD/PETab 2008. Erros-padrão robustos entre parênteses, corrigidos por correlação intra-grupo e independência inter-grupos.

Instrumentos: proibido fumar no domicílio, crença/percepção sobre as consequências do cigarro (gera doenças aos outros e para si), fumantes na família, existência de ex-fumantes na família e preço do cigarro.

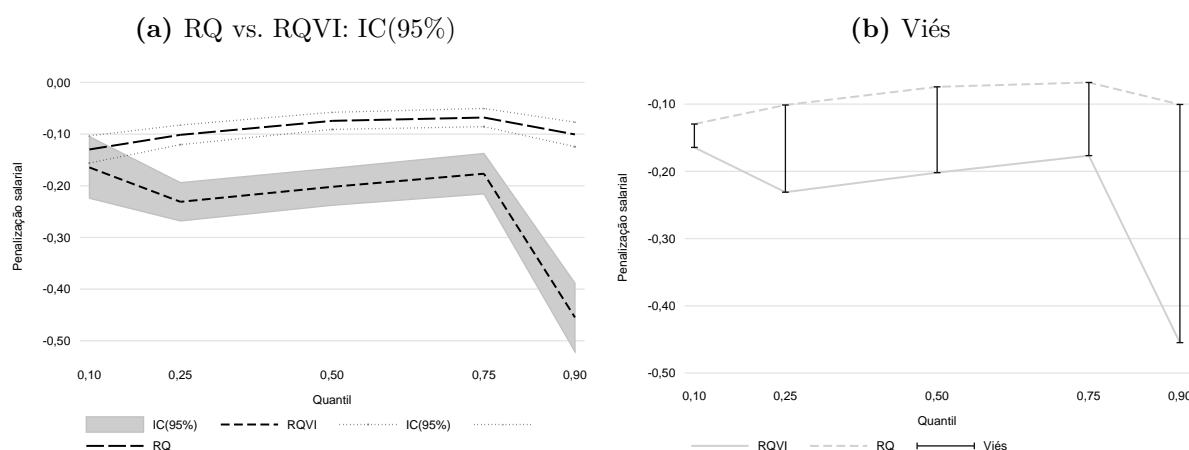
Nota: h_1 e h_2 representam os dois termos de correção do viés de seleção para estimadores quantílicos.

Legenda: *p<0,10; **p<0,05; ***p<0,01.

As variáveis de controles usuais se mostram com os sinais esperados e com significância estatística, tal como os achados para a função condicional média. Os estimadores quantílicos sem e com controle para endogeneidade da decisão de fumar sustentam a existência de um custo salarial para o uso do cigarro, independentemente da localização dos indivíduos em relação a distribuição do salário-hora. Em outras palavras, mesmo segmentando a função condicional de ganhos para trabalhadores com baixo e elevado rendimento, os fumantes indistintamente da sua posição na distribuição possui uma menor taxa de salário quando se confronta com os não-fumantes.

Através dos coeficientes obtidos pela RQ convencional, que supõem a exogeneidade da decisão de fumar, os usuários do cigarro recebem, na mediana, 7% a menos que os não-fumantes, com significância estatística a pelo menos 1%. Para os percentis 10, 25, 75 e 90, as estimativas mantêm idêntico sinal da tendência central, mas com magnitudes maiores em termos absolutos, com exceção para o quantil 0,75. Admitindo a influência dos fatores não-observados na decisão de fumar, tais estimativas, mesmo considerando os vetores de variáveis explanatórias (individuais, ocupacionais e locacionais), ainda seriam tendenciosas. Por isso, os resultados expressos na Tabela 4.7 mostram-se condizentes com as demandas técnicas requeridas nos trabalhos empíricos da área. Na intenção de facilitar a visualização e comparação dos efeitos do tabagismo sobre os salários, a Figura 4.2 exibe as estimativas por RQ e RQVI, inclusive o tamanho do viés da estimativa da penalização salarial entre as duas abordagens.

Figura 4.2 – Comparação dos efeitos do tabagismo por quantil da distribuição condicional dos rendimentos do trabalhador no Brasil entre RQ e RQVI, 2008



Fonte: Elaboração própria a partir dos microdados da PNAD/PETab 2008.

Nota: A amostra total é composta por 6.979 observações. O intervalo de confiança (IC) das estimativas considerado foi de 95%, usando erros-padrão robustos. As estimativas são reportadas para $\tau \in [0, 10; 0, 25; 0, 50; 0, 75; 0, 90]$.

Pela Figura 4.2, os custos monetários indiretos do tabagismo são maiores pela abordagem com correção para endogeneidade, sugerindo que os efeitos não corrigidos para

o problema da $Cov_\tau(H_i, \epsilon_{1i}) \neq 0$ são minorados. Resultado similar foi encontrado entre as estimativas para a função condicional média (MQO *versus* MQ2E). Enquanto o efeito por RQ varia de -12,2% a -9,6% entre os quantis 0,10 e 0,90, por RQVI estes valores variam de -15,2% a -36,5%.

Também por meio da Figura 4.2, fica evidente que a maior parte das estimativas por RQ, praticamente, não apresentam interseções com as calculadas por RQVI, mesmo considerando o intervalo de confiança (IC) de 95% – a exceção fica por parte apenas do coeficiente no quantil 0,10 entre os limites de RQ e RQVI. O viés médio entre os dois métodos é superior a 15%, com destaque para a diferença registrada no quantil 0,90. Uma das possibilidades de explicação desse resultado pode estar atrelada aos maiores níveis de aptidão técnica demandados pelas atividades com maior taxa de retorno salarial no setor privado, no qual indivíduos com maior estoque de capital humano tendem a se sobressair. Então, a junção da perda relativa de capacidade física dos fumantes, como pode ser vista na Tabela 4.3, com a desvantagem da formação do capital humano motivado por menores investimentos sistemáticos no estoque de saúde são prováveis fatores explicativos.

Além do tabagismo ser um dos principais causadores de mortes dentre os fatores de risco modificáveis (WHO, 2009), os usuários do cigarro no Brasil, conforme os presentes resultados, arcam também com elevados custos monetários indiretos. Com evidências semelhantes para a Finlândia, Bockerman, Hyytinen & Kaprio (2014) questionam até que ponto os usuários do cigarro conseguem calcular de forma racional a relação de custo-benefício do produto como admitido em Becker & Murphy (1988).

4.5.2 Diferenciação compensatória salarial para atividades perigosas

Esta seção testa a proposição do modelo de Viscusi & Hersch (2001) de diferenciais compensatórios de salários entre fumantes e não-fumantes, complementando os achados sobre os custos monetários indiretos do tabagismo, expostos anteriormente, ao examinar a relação desse grupo de trabalhadores com as atividades de maior periculosidade de acidentes.

A Tabela 4.8 exibe as estimativas para o parâmetro a_1^s relativo aos riscos de acidentes – que representa o prêmio ao risco ($w_p = \partial w / \partial p$) –, com especificação econométrica aplicada separadamente por status de fumante para modelos sem e com VI⁵⁹.

⁵⁹ Como exibido em todas as tabelas na parte de penalização, os coeficientes das variáveis de controle (tais como escolaridade, raça, idade, fatores locais etc.) mostram-se com idênticos sinais e em sua maioria estatisticamente significativos, por isso a supressão deles nesta parte.

Tabela 4.8 – Coeficiente do prêmio ao risco de acidentes do trabalho e VSI no Brasil por status de fumante e diferentes modelagens econométricas, 2008

Detalhes	Média	Mediana	Q(0,10)	Q(0,25)	Q(0,75)	Q(0,90)
Não-fumantes						
Sem VI						
Coeficiente (w_p)	0,0491***	0,0514***	0,0473***	0,0527***	0,0423***	0,0362***
Erro-padrão	0,004	0,003	0,0046	0,0037	0,0047	0,0054
VSI (em R\$ de 2008)	2.094,99	2.099,71	1.083,64	1.570,15	2.473,66	3.040,03
VSI (em US\$ de 2000)	662,77	664,26	342,82	496,73	782,57	961,74
R ² ajustado/Pseudo R ²	0,46	0,27	0,17	0,22	0,32	0,34
Com VI						
Coeficiente (w_p)	0,1357***	0,1677***	0,1312***	0,1390***	0,2156***	0,1960***
Erro-padrão	0,0105	0,0128	0,0106	0,0106	0,023	0,031
VSI (em R\$ de 2008)	5.787,59	6.389,54	2.894,51	3.923,26	12.297,47	16.738,09
VSI (em US\$ de 2000)	1.830,96	2.021,40	915,7073	1.241,16	3.890,43	5.295,27
R ² ajustado	0,44	n/d	n/d	n/d	n/d	n/d
N	22.088	22.088	22.088	22.088	22.088	22.088
P	10.739.467	10.739.467	10.739.467	10.739.467	10.739.467	10.739.467
Fumantes						
Sem VI						
Coeficiente (w_p)	0,0335***	0,0349***	0,0258***	0,0316***	0,0197***	0,0266**
Erro-padrão	0,0066	0,0069	0,0057	0,0059	0,0035	0,0114
VSI (em R\$ de 2008)	1.153,24	1.167,40	485,93	779,56	920,04	1.727,73
VSI (em US\$ de 2000)	364,84	369,32	153,73	246,62	291,06	546,59
R ² ajustado/Pseudo R ²	0,42	0,22	0,16	0,17	0,28	0,33
Com VI						
Coeficiente (w_p)	0,0813***	0,0967***	0,0548**	0,0857***	0,1007***	0,1335**
Erro-padrão	0,0223	0,0251	0,0259	0,0235	0,0368	0,0536
VSI (em R\$ de 2008)	2.799,24	3.542,08	1.118,45	2.380,30	5.052,87	9.386,27
VSI (em US\$ de 2000)	885,57	1.120,57	353,83	753,03	1.598,53	2.969,44
R ² ajustado	0,42	n/d	n/d	n/d	n/d	n/d
N	8.153	8.153	8.153	8.153	8.153	8.153
P	4.080.790	4.080.790	4.080.790	4.080.790	4.080.790	4.080.790
Atributos individuais	X	X	X	X	X	X
Fatores locais	X	X	X	X	X	X
Controle: viés de seleção	X	X	X	X	X	X
Controle: ex-fumantes	X	X	X	X	X	X

Fonte: Elaboração própria a partir dos microdados da PNAD/PETab 2008 e do AEAT/MPS 2008. Erros-padrão robustos, corrigidos por correlação intra-grupo e independência inter-grupos. n/d = não disponível.

Instrumentos: renda total do não-trabalho, TOA (em nível e ao quadrado), estrutura familiar (casado sem filhos e casado com filhos), chefe do domicílio, filho com dificuldades físicas e cônjuge com dificuldades físicas.

Nota: Os trabalhadores da agricultura e conta-própria foram desconsiderados da amostra para fins comparativos com os resultados de Viscusi & Hersch (2001). VSI computado supondo 2.000 horas de trabalho para anualização dos salários previstos. Para a obtenção do VSI para dólares do ano 2000, fez-se a conversão dos valores em reais de 2008 para o ano 2000 por meio do Índice Nacional de Preços ao Consumidor Amplo (IPCA) e, em seguida, utilizou-se a taxa de câmbio real (R\$) por dólar (US\$) do ano 2000 disponível na plataforma *Penn World Table* (<http://pwt.sas.upenn.edu>) de Feenstra, Inklaar & Timmer (2013).

Legenda: *p<0,10; **p<0,05; ***p<0,01.

O pressuposto de diferenciais compensatórios salariais é verificado se os coeficientes do prêmio ao risco de acidentes não-fatais do trabalho (w_p) forem significativos e maiores do que zero. Segundo a Tabela 4.8, o parâmetro α_1^s apresenta sinal positivo e estatisticamente significativo a pelo menos 5% de significância em todos os modelos para os dois grupos de

trabalhadores. Muito embora, as abordagens com variáveis instrumentais se diferenciam pela maior magnitude dos coeficientes.

Essas evidências, acompanhando os trabalhos de [Smith \(1979\)](#), [Viscusi & Moore \(1987\)](#), [Viscusi & Hersch \(2001\)](#), [Viscusi & Aldy \(2003\)](#), [Esteves \(2008\)](#), [Fernandez & Nordman \(2009\)](#), [Evans & Schaur \(2010\)](#) e [Contzen, Won & Lavin \(2013\)](#), sugerem que trabalhadores pertencentes às atividades econômicas com maiores probabilidades de risco de acidentes do trabalho possuem um salário-hora superior aos de trabalhadores de atividades menos arriscadas, mesmo admitindo uma série de controles observáveis individuais, correção de viés de seleção amostral e diferentes modelos econométricos. No Painel B da Tabela C.4 no Apêndice estão disponíveis outras especificações de instrumentos e indicadores do ajustamento para o primeiro estágio, cujo os valores certificam as tendências expostas na Tabela 4.8 e, em termos globais, qualificam os instrumentos utilizados⁶⁰.

Como preconizado no modelo e nos achados empíricos de [Viscusi & Hersch \(2001\)](#), os diferenciais compensatórios requeridos pelos trabalhadores não-fumantes são superiores, em quaisquer especificações, aos constatados para os usuários do cigarro. No modelo elaborado para a função condicional média com variáveis instrumentais (MQ2E), os não-fumantes e os fumantes dispõem, respectivamente, de um coeficiente de 0,136 e 0,081. Já por meio dos estimadores quantílicos, relativamente, as maiores diferenças entre os coeficientes para os dois grupos de trabalhadores são verificadas nos quantis 0,10 e 0,75, onde as maiores taxas de compensação requerida para ambos os grupos de trabalhadores ocorrem no *locus* da distribuição acima da mediana (0,2156 no quantil 0,75 para os não-fumantes e 0,1335 no quantil 0,90 para os fumantes).

Esses resultados se aproximam do *trade-off* salário-risco exemplificado na Figura 4.1 para o caso em que $w_s < 0$ e $w_{ps} < 0$, pelos seguintes pontos: através dos resultados da penalização salarial do tabagismo (visto na seção anterior), os trabalhadores fumantes recebem uma taxa de salário inferior aos dos não-usuários, implicando um efeito negativo do consumo do cigarro sobre os salários; e, em termos comparativos, a diferença entre os parâmetros a_1^f e a_1^{nf} é menor do que zero, significando em menor prêmio ao risco recebido pelos os usuários do cigarro. Com isso, os fumantes tendem a ter uma curva de oferta menos inclinada do que a dos não-fumantes, de maneira que para um mesmo indicador de atividade perigosa, os fumantes exigem relativamente uma menor compensação salarial, o que na linguagem de [Hersch & Viscusi \(1990\)](#) representa uma menor aversão revelada ao risco por parte desse grupo de indivíduos. Portanto, os trabalhadores não-fumantes escolhem trabalhos mais perigosos se tiverem uma maior compensação de salários, enquanto que os fumantes mostram-se menos avessos ao risco, por suportarem os perigos de acidentes

⁶⁰ Pelos testes de endogeneidade de Durbin e Wu-Hausman, a hipótese de exogeneidade para o índice de risco (IRAT) é rejeitada em ambos os testes, haja vista o p-valor aproximadamente igual a zero para os escores χ^2 e F dos testes para a regressão dos salários hedônicos para os não-fumantes e fumantes.

do trabalho por um menor nível de prêmio.

Outrossim, a tendência observada pelos coeficientes de diferenciação compensatória salarial ao risco de acidentes por RQVI é, em geral, crescente em relação ao quantil condicional dos rendimentos tanto para os fumantes quanto para os não-fumantes. Logo, trabalhadores situados em pontos mais elevados da distribuição condicional do salário-hora, que propendem a ter maior produtividade do trabalho, exigem uma maior taxa de compensação salarial para atividades com maior periculosidade. Assim como em [Fernandez & Nordman \(2009\)](#) e [Evans & Schaur \(2010\)](#), esta pesquisa detecta prêmios ao risco com características heterogêneas dependendo do quantil da distribuição salarial.

Dadas as estimativas para a compensação salarial para o risco de acidentes não-fatais, pode-se computar o VSI para o caso brasileiro (ver na Tabela 4.8) a partir, por exemplo, de [Viscusi & Moore \(1987\)](#), [Viscusi & Hersch \(2001\)](#) e [Viscusi & Aldy \(2003\)](#). Vale lembrar, que o coeficiente de prêmio ao risco estimado corresponde a uma parte do salário relativo em que os trabalhadores estão dispostos a aceitar a fim de suportar o risco de uma atividade com chances de acidentes não-fatais. Assim, o valor estatístico de um acidente (VSI) é obtido com base nessa interpretação e na média do salário anual previsto para cada grupo de trabalhadores. Na mediana o VSI para não-fumantes é de R\$ 6.389,54 e para os consumidores do cigarro de R\$ 3.542,08 por acidente. Quando se verifica os indivíduos mais produtivos (quantis 0,75 e 0,90) que nunca usaram cigarro, o valor estatístico de um acidente é superior a 12 mil reais, ao tempo que os trabalhadores localizados abaixo da mediana da distribuição recebem menos de R\$ 4 mil. Para os fumantes, os valores estimados pelos quantis mais baixos e mais altos são, na média, a metade do VSI recebido pelo o outro grupo. A título de comparação, o VSI estimado por [Viscusi & Hersch \(2001\)](#) dos fumantes nos EUA equivale a 44% do calculado para os não-fumantes.

O VSI para o mercado de trabalho norte-americano, medido em dólares do ano 2000, entre os mais de 30 estudos sobre o tema, segundo [Viscusi & Aldy \(2003\)](#), varia no intervalo de US\$ 20.000 a US\$70.000 anuais por acidente, dependendo do método, do período e da variável *proxy* para o risco de acidente. Restringindo para os trabalhos que supõem percepções ao risco variáveis, os fumantes têm prêmios ao risco de acidentes mais baixos do que os não-fumantes, correspondente ao encontrado pela presente pesquisa para o caso brasileiro. Em [Viscusi & Hersch \(2001\)](#), os não-fumantes exibem um VSI de US\$ 47.476 e os fumantes de US\$ 20.755 – os dois valores atualizados para dólares de 2000 –, considerando a covariada do BLS *Injury Rate* por 100 trabalhadores e um modelo de regressão condicionado para a média. No confronto direto com o citado trabalho, o VSI estimado para a tendência central da distribuição condicional do salário para o Brasil, medido em dólares de 2000, é bem inferior ao do caso norte-americano. Além disso, o montante estimado do VSI brasileiro para os quantis mais altos da distribuição condicional de salários, que capta os indivíduos mais produtivos no país, são muito menores do que o

calculado por [Viscusi & Hersch \(2001\)](#) entre fumantes e não-fumantes com produtividade média nos EUA.

4.6 Conclusões

As evidências do presente estudo suportam a existência de uma penalização salarial do tabagismo no mercado de trabalho brasileiro, que se sustentam mesmo após uma série de controles para endogeneidade, amostra, atributos observáveis individuais e posição do trabalhador na distribuição condicional dos rendimentos. Usando diferentes instrumentais econométricos, com ênfase no modelo RQVI, o usuário do tabaco no Brasil arca com um custo monetário indireto do produto, percebido pelos menores níveis da taxa de salário-hora. Complementando a literatura da área, as estimativas indicam que as penalizações salariais não se mostram constantes em diferentes partes da distribuição dos ganhos. Assim, as perdas, controlando para endogeneidade, variam de 15,2% a 36,5%, centrando em torno de 18,3% na mediana e alcançando 36,5% para os trabalhadores no extremo positivo da distribuição salarial.

Além do mais, a exemplo dos resultados de [Hersch & Viscusi \(1990\)](#) e [Viscusi & Hersch \(2001\)](#), os trabalhadores brasileiros usuários do cigarro recebem relativamente uma menor compensação salarial para atividades insalubres no comparativo com os não-usuários do produto. As regressões sem controle para endogeneidade reportam a existência de um prêmio ao risco de acidentes do trabalho não-fatais, mas com menor magnitude do que o auferido com o método com VI. Com base no modelo RQVI, a estimativa para o valor estatístico de um acidente do trabalho reforça essas diferenças entre os dois grupos, com uma diferença em favor dos não-usuários do cigarro de, aproximadamente, R\$ 1,78 mil e R\$ 7,35 mil, respectivamente, nos quantis 0,10 e 0,90 do valor estatístico de um acidente anualizado. Não obstante, o VSI para o caso nacional é muito inferior, por exemplo, ao identificado por [Viscusi & Hersch \(2001\)](#) para o mercado de trabalho dos EUA, já que comparativamente o valor esperado corresponde a 4,3% para os não-fumantes e a 5,4% para os fumantes do VSI norte-americano, aumentando, na ordem, para 11,2% e 14,3% para os indivíduos brasileiros com maior produtividade marginal.

A combinação de uma elevada penalização salarial com um menor nível de compensação ao risco mostra que as sequelas do tabagismo vão, em termos individuais, além das questões de saúde por si só, atingindo custos monetários implícitos que podem repercutir em médio e longo prazo, inclusive, em um menor patamar de recurso disponível para remédios e tratamentos de doenças decorrentes das próprias substâncias maléficas do produto ao organismo. Assim, políticas anti-tabagismo, sobretudo as preventivas, podem gerar impactos não apenas nas condições de saúde, como redução da taxa de mortalidade e aumento no tempo de sobrevida, mas também em indicadores econômicos no prisma

individual e social.

Por fim, este estudo faz uma contribuição importante para a literatura nacional, pois é um dos primeiros a relacionar o tabagismo e os resultados no mercado de trabalho no Brasil, bem como incorpora a heterogeneidade dos efeitos do uso do cigarro nos salários e no cálculo do prêmio ao risco em consonância com a posição do trabalhador na distribuição condicional dos rendimentos. Embora os dados utilizados não sejam longitudinais (fato ainda recorrente nas pesquisas estatísticas nacionais), esse tipo de avaliação dos custos e das consequências do uso de substâncias nocivas à saúde deveriam ser mais enfatizadas no país, a fim de melhorar a alocação de recursos para combate aos problemas relacionados ao uso do cigarro e melhor compreender as relações envolvendo escolhas não saudáveis e resultados econômicos e não-econômicos.

5 Considerações Finais

Nos últimos anos, o poder público brasileiro vem demonstrando esforços para ampliar a cobertura e a qualidade dos serviços providos pelo o Estado nas áreas de saúde e de educação⁶¹, bem como na criação de políticas de controle tarifário e punitivo para inibição do uso de produtos de risco⁶², de sistemas de avaliação e controle nessas áreas com ampliação da disseminação de informações estatísticas⁶³ – que são insumos importantes para pesquisas acadêmicas e subsídio para diagnóstico e elaboração de políticas públicas. Mesmo supondo a existência de um vasto serviço público em termos de atendimento e qualidade, as preferências individuais dos próprios agentes, como a adoção de bons hábitos, continuariam sendo aspectos decisivos para manutenção ou ampliação do estoque de capital humano, como ressaltam Grossman (1972) e Cawley & Ruhm (2011).

Diante das evidências encontradas, a adoção de hábitos prejudiciais à saúde provoca consideráveis penalizações para a formação do capital humano do país. Em termos dos resultados educacionais, as crianças expostas aos fatores de risco à saúde possuem um menor desempenho escolar, sobretudo as de menor nível socioeconômico. Esse cenário revela também a importância dos efeitos dos pares (dos colegas de escola e familiares) na determinação das escolhas não saudáveis da criança. Logo, é de fundamental interesse para as políticas públicas a identificação dos grupos mais susceptíveis a iniciação do consumo de produtos não saudáveis ou aos fatores de risco modificáveis à saúde para reduzir as probabilidades de perpetuação a médio e longo prazo de agentes com baixo estoque de capital humano e, conseqüentemente, menores taxas de retornos econômicas e não-econômicas. A alta penalização salarial do tabagismo no Brasil é um exemplo claro dos efeitos adversos de posturas não saudáveis sobre a produtividade marginal do trabalho, adicionalmente os menores valores estimados para o VSI dos fumantes quando comparados ao recebido pelos não-fumantes demonstram que os usuários do cigarro são menos avessos ao risco de acidentes do trabalho.

Com base na análise da demanda, o consumo de bebidas alcoólicas é elástico a alteração na renda familiar, inclusive para as famílias público-alvo dos programas sociais de transferência de renda, ao passo que o cigarro é classificado como produto necessário.

⁶¹ Esses esforços do poder público podem ser sinalizados pelo o aumento em mais de 36% dos gastos públicos com saúde e educação como proporção do PIB entre 2000 e 2009, conforme os dados da plataforma WDI do Banco Mundial.

⁶² Como a lei nº 11.705/2008 do Código de Trânsito Brasileiro, conhecida como Lei Seca, e da elevação de alíquotas tributárias, por exemplo, para a cerveja expressa na Portaria do Ministério da Fazenda nº 181 de 31/03/2014.

⁶³ Censo da Educação Básica, Censo da Educação Superior, Enem, Prova Brasil e Enade são exemplos de microdados na área de educação, já para saúde o sistema Datasus disponibiliza informações detalhadas sobre morbidade, mortalidade, nascidos vivos, procedimentos ambulatoriais e hospitalares.

Ademais, o baixo ajustamento da demanda no cenário de choques dos preços do cigarro e bebidas alcoólicas e a substitutibilidade dos dois produtos em termos de preços-cruzados, indicam que políticas tarifárias sobre os citados itens geram potenciais efeitos adversos no consumo de itens alimentares e a taxaço em apenas um dos produtos, por exemplo, não se reflete em uma redução conjunta na demanda do outro item. Assim, as políticas não-tarifárias, focadas na prevenção, em princípio se apresentariam como uma das melhores alternativas para atenuar, especialmente a exposiço de crianças e adolescentes para não reproduzir baixos resultados na escola e também no mercado de trabalho.

Referências

- ABREVAYA, J.; TANG, H. Body mass index in families: spousal correlation, endogeneity, and intergenerational transmission. *Empirical Economics*, v. 41, n. 3, p. 841–864, out. 2010. ISSN 0377-7332.
- ALBERNAZ, A.; FERREIRA, F.; FRANCO, C. Qualidade e Eqüidade no Ensino Fundamental Brasileiro. *Pesquisa e Planejamento Econômico*, v. 32, n. 3, p. 453–476, 2002.
- ALMEIDA, A. T. C. Determinantes dos piores e melhores resultados educacionais dos alunos da rede pública de ensino fundamental no Brasil. *Planejamento e Políticas Públicas*, v. 42, p. 147–188, 2014.
- ALVES, L. F.; ANDRADE, M. V. Impactos da saúde nos rendimentos individuais no Brasil. *Economia Aplicada*, v. 7, n. 2, p. 259–388, 2003.
- ALVES, M. T. G.; SOARES, J. F. *O nível socioeconômico das escolas de educação básica brasileiras*. Belo Horizonte: UFMG, 2012. 1–57 p.
- ANGER, S.; KVASNICKA, M. Does Smoking Really Harm Your Earnings so Much? Biases in Current Estimates of the Smoking Wage Penalty. *Applied Economics Letters*, v. 17, n. 6, p. 561–564, 2010.
- ARNOULD, R. J.; NICHOLS, L. M. Wage-Risk Premiums and Workers ' Compensation: A Refinement of Estimates of Compensating Wage Differential. *Journal of Political Economy*, v. 91, n. 2, p. 332–340, 1983.
- AULD, M. C. Smoking, Drinking, and Income. *The Journal of Human Resources*, v. 40, n. 2, p. 505–518, 2005.
- BALSA, A. I.; GIULIANO, L. M.; FRENCH, M. T. The effects of alcohol use on academic achievement in high school. *Economics of Education Review*, Elsevier Ltd, v. 30, n. 1, p. 1–15, fev. 2011. ISSN 0272-7757.
- BANKS, J.; BLUNDELL, R.; LEWBEL, A. Quadratic Engel Curves and Consumer Demand. *The Review of Economics and Statistics*, v. 79, n. 4, p. 527–539, 1997.
- BARROS, R. P.; MENDONÇA, R.; SANTOS, D. D.; QUINTAES, G. Determinantes do desempenho educacional no Brasil. *Pesquisa e Planejamento Econômico*, v. 31, n. 1, p. 1–42, 2001.
- BARTEL, A.; TAUBMAN, P. Health and Labor Market Success: The Role of Various Diseases. *The Review of Economics and Statistics*, v. 61, n. 1, p. 1–8, 1979.
- BARTEN, A. P. Demand Functions under Conditions of Almost Additive Preferences. *Econometrica*, v. 32, n. 1, p. 1–38, 1964.
- _____. Maximum likelihood estimation of a complete system of demand equations. *European Economic Review*, v. 1, n. 1, p. 7–73, 1969.

- BECKER, G. S.; MULLIGAN, C. B. The Endogenous Determination of Time Preference. *The Quarterly Journal of Economics*, v. 112, n. 3, p. 729–758, 1997.
- BECKER, G. S.; MURPHY, K. M. A Theory of Rational Addiction. *Journal of Political Economy*, v. 96, n. 4, p. 675–700, 1988.
- BILGIC, A.; YEN, S. T. Household food demand in Turkey: A two-step demand system approach. *Food Policy*, Elsevier Ltd, v. 43, p. 267–277, dez. 2013. ISSN 03069192.
- BLACKLOW, P.; NICHOLAS, A.; RAY, R. Demographic Demand Systems With Application To Equivalence Scales Estimation and Inequality Analysis: the Australian Evidence. *Australian Economic Papers*, v. 49, n. 3, p. 161–179, ago. 2010. ISSN 0004900X.
- BLUNDELL, R.; PASHARDES, P.; WEBER, G. What Do We Learn About Consumer Demand Patterns from Micro Data? *The American Economic Review*, v. 83, n. 3, p. 570–597, 1993.
- BLUNDELL, R.; ROBIN, J. M. Estimation in large and disaggregated Demand Systems: an estimator for conditionally Linear Systems. *Journal of Applied Econometrics*, v. 14, n. 3, p. 209–232, 1999.
- BOCKERMAN, P.; HYYTINEN, A.; KAPRIO, J. Smoking and long-term labour market outcomes. *Tobacco control*, p. 1–7, fev. 2014. ISSN 1468-3318. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/24570102>>.
- BRAAKMANN, N. The smoking wage penalty in the United Kingdom: Regression and matching evidence from the British Household Panel Survey. *University of Lüneburg Working Paper Series N. 96*, p. 1–33, 2008.
- BRASIL. *A vigilância , o controle e a prevenção das doenças crônicas não transmissíveis DCNT no contexto do sistema único de saúde brasileiro: situação e desafios atuais*. Brasília-DF: Ministério da Saúde, 2005.
- _____. *Plano de Ações Estratégicas para o enfrentamento das doenças crônicas não transmissíveis (DCNT) no Brasil: 2011-2022*. Brasília-DF: Ministério da Saúde, 2011. ISBN 9788533418318.
- BUCHINSKY, M. The dynamics of changes in the female wage distribution in the USA: a quantile regression approach. *Journal of Applied Econometrics*, v. 13, n. 1, p. 1–30, 1998.
- _____. Quantile regression with sample selection: Estimating women's return to education in the U.S. *Empirical Economics*, v. 26, n. 1, p. 87–113, 2001.
- CALIENDO, M.; KOPEINIG, S. Some Practical Guidance for the Implementation of Propensity Score Matching. *IZA Working Paper n. 1588*, p. 1–32, 2005.
- CARRELL, S. E.; HOEKSTRA, M.; WEST, J. E. Does drinking impair college performance? Evidence from a regression discontinuity approach. *Journal of Public Economics*, Elsevier B.V., v. 95, n. 1-2, p. 54–62, fev. 2011. ISSN 00472727. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0047272710001143>>.
- CARVALHO, D. B.; SIQUEIRA, R. B. D.; NOGUEIRA, R. B. Características Distributivas e Impacto de Reformas Tributárias Sobre o Bem-Estar das Famílias no Brasil. *Revista Brasileira de Economia*, v. 67, n. 3, p. 263–282, 2013.

- CARVALHO, J. L.; LOBAO, W. Vício privado e políticas públicas: a demanda por cigarros no Brasil. *Revista Brasileira de Economia*, v. 52, p. 67–104, 1998.
- CAWLEY, J.; RUHM, C. The Economics of Risky Health Behaviors. *NBER Working Paper n. 17081*, p. 1–162, 2011. Disponível em: <<http://www.nber.org/papers/w17081>>.
- CHALOUPKA, F. Rational Addictive Behavior and Cigarette Smoking. *Journal of Political Economy*, v. 99, n. 4, p. 722–742, 1991.
- CHALOUPKA, F.; WARNER, K. E. The economics of smoking. *NBER Working Paper n. 7047*, p. 1–70, 1999.
- CHALOUPKA, F. J.; GROSSMAN, M. Price, Tobacco Control Policies and Youth Smoking. *NBER Working Paper n. 5740*, p. 1–41, 1996. Disponível em: <<http://www.nber.org/papers/w5740.pdf>>.
- CHATTERJI, P. Does alcohol use during high school affect educational attainment?: Evidence from the National Education Longitudinal Study. *Economics of Education Review*, v. 25, n. 5, p. 482–497, out. 2006. ISSN 02727757. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0272775705000774>>.
- CHERNOZHUKOV, V.; HANSEN, C. The Effects of 401(k) Participation on the Wealth Distribution: An Instrumental Quantile Regression Analysis. *The Review of Economics and Statistics*, v. 86, n. 3, p. 735–751, 2004.
- _____. Notes and comments an IV model of quantile treatment effects. *Econometrica*, v. 73, n. 1, p. 245–261, 2005.
- _____. Quantile Models with Endogeneity. *Annual Review of Economics*, v. 5, n. 1, p. 57–81, jan. 2013. ISSN 1941-1383. Disponível em: <<http://www.annualreviews.org/doi/abs/10.1146/annurev-economics-080511-110952>>.
- CHERNOZHUKOV, V.; HANSEN, C.; JANSSEN, M. Inference approaches for instrumental variable quantile regression. *Economics Letters*, v. 95, n. 2, p. 272–277, maio 2007. ISSN 01651765. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0165176506003703>>.
- CHRISTENSEN, L. R.; JORGENSON, D. W.; LAU, L. J. Transcendental Logarithmic Utility Functions. *The American Economic Review*, v. 65, n. 3, p. 367–383, 1975.
- COLEMAN, J. S.; HOPKINS, J.; CAMPBELL, E. Q.; HOBSON, C. J. *Equality of educational opportunity*. Washington, D.C.: U.S. Government Printing Office, 1966. 1–737 p.
- CONTI, G.; HECKMAN, J.; URZUA, S. The Education-Health Gradient. *American Economic Review*, v. 100, n. 2, p. 234–238, maio 2010. ISSN 0002-8282.
- CONTZEN, M. P.; WON, A. R.; LAVIN, F. V. The value of a statistical life in Chile. *Empirical Economics*, v. 45, n. 3, p. 1073–1087, nov. 2013. ISSN 0377-7332. Disponível em: <<http://link.springer.com/10.1007/s00181-012-0660-7>>.
- CORNFIELD, J.; HAENSZEL, W.; HAMMOND, E. C.; LILIENFELD, A. M.; SHIMKIN, M. B.; WYNDER, E. L. Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions. *Jornal of the National Cancer Institute*, v. 22, n. 1, p. 173–203, out. 1959. ISSN 1464-3685.

- CROSNOE, R.; MULLER, C. Body Mass Index, Academic Achievement, and School Context: Examining the Educational Experiences of Adolescents at Risk of Obesity. *Journal of Health and Social Behavior*, v. 45, n. 4, p. 393–407, dez. 2004. ISSN 0022-1465.
- CUTLER, D.; GLAESER, E. What explains differences in smoking, drinking and other health-related behaviors? *NBER Working paper series n. 11100*, p. 1–20, 2005.
- CUTLER, D. M.; LLERAS-MUNEY, A. Understanding differences in health behaviors by education. *Journal of Health Economics*, v. 29, n. 1, p. 1–28, jan. 2010. ISSN 1879-1646.
- DE LUCA, G. The Stata Journal. *The Stata Journal*, v. 8, n. 2, p. 190–220, 2008.
- DEATON, A.; MUELLBAUER, J. An Almost Ideal Demand System. *The American Economic Review*, v. 70, n. 3, p. 312–326, 1980.
- DECKER, S. L.; SCHWARTZ, A. E. Cigarettes and Alcohol: substitutes or complements? *NBER Working Paper n. 7535*, 2000.
- DEOLALIKAR, A. B. Nutrition and Labor Productivity in Agriculture: Estimates for Rural South India. *The Review of Economics and Statistics*, v. 70, n. 3, p. 406–413, 1988.
- DING, W.; LEHRER, S. F.; ROSENQUIST, J. N.; AUDRAIN-MCGOVERN, J. The impact of poor health on academic performance: New evidence using genetic markers. *Journal of Health Economics*, v. 28, n. 3, p. 578–97, maio 2009. ISSN 0167-6296.
- DIPRETE, T. A.; GANGL, M. Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments. *Sociological Methodology*, v. 34, n. 1, p. 271–310, 2004.
- DUFFY, M. Advertising in Demand Systems for Alcoholic Drinks and Tobacco: A Comparative Study. *Journal of Policy Modeling*, v. 17, n. 6, p. 557–577, 1995.
- _____. Advertising and food, drink and tobacco consumption in the United Kingdom: a dynamic demand system. *Agricultural Economics*, v. 28, n. 1, p. 51–70, jan. 2003. ISSN 01695150.
- EAKINS, J. M.; GALLAGHER, L. A. Dynamic almost ideal demand systems: an empirical analysis of alcohol expenditure in Ireland. *Applied Economics*, v. 35, n. 9, p. 1025–1036, 2003.
- EIDE, E.; SHOWALTER, M. The effect of school quality on student performance: A quantile regression approach. *Economics Letters*, v. 58, p. 345–350, 1998.
- EIDE, E. R.; SHOWALTER, M. H. Estimating the relation between health and education: What do we know and what do we need to know? *Economics of Education Review*, Elsevier Ltd, v. 30, n. 5, p. 778–791, out. 2011. ISSN 02727757. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0272775711000525>>.
- EIDE, E. R.; SHOWALTER, M. H.; GOLDBABER, D. D. The relation between children's health and academic achievement. *Children and Youth Services Review*, Elsevier Ltd, v. 32, n. 2, p. 231–238, fev. 2010. ISSN 01907409.

ELLICKSON, P. L.; TUCKER, J. S.; KLEIN, D. J. High-Risk Behaviors Associated With Early Smoking: Results From a 5-Year Follow-Up. *Journal of Adolescent Health*, v. 28, n. 6, p. 465–473, 2001.

ESTEVEES, L. A. Salários e risco de acidentes do trabalho: evidências de diferenciais compensatórios para a indústria manufatureira. *Economia Aplicada*, v. 12, n. 2, p. 275–287, 2008.

EVANS, M. F.; SCHAUR, G. A quantile estimation approach to identify income and age variation in the value of a statistical life. *Journal of Environmental Economics and Management*, Elsevier, v. 59, n. 3, p. 260–270, maio 2010. ISSN 00950696. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0095069609001077>>.

FAN, S.; CRAMER, G.; WAILES, E. Food demand in rural China: evidence from rural household survey. *Agricultural Economics*, v. 11, p. 61–69, 1994.

FEENSTRA, R. C.; INKLAAR, R.; TIMMER, M. *The Next Generation of the Penn World Table*. 2013. 1–42 p. Disponível em: <www.ggdc.net/pwt>.

FERNANDEZ, R. M.; NORDMAN, C. J. Are there pecuniary compensations for working conditions? *Labour Economics*, Elsevier B.V., v. 16, n. 2, p. 194–207, abr. 2009. ISSN 09275371. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0927537108000791>>.

FLORIN, T. A.; SHULTS, J.; STETTLER, N. Perception of overweight is associated with poor academic performance in US adolescents. *The Journal of School Health*, v. 81, n. 11, p. 663–70, nov. 2011. ISSN 1746-1561.

FRIEDMAN, J.; LEVINSOHN, J. The Distributional Impacts of Indonesia's Financial Crisis on Household Welfare: A "Rapid Response" Methodology. *The World Bank Economic Review*, v. 16, n. 3, p. 397–423, dez. 2002. ISSN 1564698X.

GALLANT, A. R.; NYCHKA, D. W. Semi-Nonparametric Maximum Likelihood Estimation. *Econometrica*, v. 55, n. 2, p. 363–390, 1987.

GALLET, C. a.; LIST, J. a. Cigarette demand: a meta-analysis of elasticities. *Health Economics*, v. 12, n. 10, p. 821–35, out. 2003. ISSN 1057-9230. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/14508867>>.

GAREN, J. Compensating Wage Differentials and the Endogeneity of Job Riskiness. *The Review of Economics and Statistics*, v. 70, n. 1, p. 9–16, 1988.

GASTWIRTH, J. L.; KRIEGER, A. M.; ROSENBAUM, P. R. Dual and simultaneous sensitivity analysis for matched pairs. *Biometrika*, v. 85, n. 4, p. 907–920, 1998.

GIL, A. I.; MOLINA, J. A. Alcohol demand among young people in Spain: an addictive QUAIDS. *Empirical Economics*, v. 36, n. 3, p. 515–530, jun. 2009. ISSN 0377-7332.

GLEWWE, P.; JACOBY, H. G. An economic analysis of delayed primary school enrollment in a low income country: the role of early childhood nutrition. *The Review of Economics and Statistics*, v. 77, n. 1, p. 156–169, 1995.

GLEWWE, P. W.; HANUSHEK, E. A.; HUMPAGE, S. D.; RAVINA, R. School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010. *NBER Working paper series n. 17554*, p. 1–71, 2011. Disponível em: <http://www.nber.org/papers/w17554>.

GOEL, R. K.; MOREY, M. J. The Interdependence of Cigarette and Liquor Demand. *Southern Economic Journal*, v. 62, n. 2, p. 451–459, 1995.

GOMES-NETO, J. B.; HANUSHEK, E. A. Causes and Consequences of Grade Repetition: Evidence from Brazil. *Economic Development and Cultural Change*, v. 43, n. 1, p. 117–148, 1994.

GOMES-NETO, J. B.; HANUSHEK, E. A.; LEITE, R. H.; FROTA-BEZZERA, R. C. Health and Schooling: Evidence and Policy Implications for Developing Countries. *Economics of Education Review*, v. 16, n. 3, p. 271–282, 1997.

GRAFOVA, I. B.; STAFFORD, F. P. The Wage Effects of Personal Smoking History. *Industrial and Labor Relations Review*, v. 62, n. 3, p. 381, abr. 2009. ISSN 0019-7939.

GROSSMAN, M. On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy*, v. 80, n. 2, p. 223–255, 1972.

HALVORSEN, R.; PALMQUIST, R. The Interpretation of Dummy Variables in Semilogarithmic Equations. *The American Economic Review*, v. 70, n. 3, p. 474–475, 1980.

HANUSHEK, E. A. The Production of Education, Teacher Quality and Efficiency. In: U.S. Office of Education (Ed.). *Do teachers make a difference?* Washington, D.C.: Government Printing Office, 1970. cap. 4, p. 79–99.

_____. Teacher characteristics and gains in student achievement: Estimation using micro data. *The American Economic Review*, v. 61, n. 2, p. 280–288, 1971. Disponível em: <http://www.jstor.org/stable/10.2307/1817003>.

_____. The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature*, v. 24, n. 3, p. 1141–1177, 1986.

HECKMAN, J. J. Sample Selection Bias as a Specification Error. *Econometrica*, v. 47, n. 1, p. 153–161, 1979.

HEIEN, D.; WESSELLS, C. R. Demand Systems Estimation with Microdata: A Censored Regression Approach. *Journal of Business & Economic Statistics*, v. 8, n. 3, p. 365–371, 1990.

HERSCH, J.; VISCUSI, W. K. Cigarette Smoking, Seatbelt Use, and Differences in Wage-Risk Tradeoffs. *The Journal of Human Resources*, v. 25, n. 2, p. 202–227, 1990.

Instituto Brasileiro de Geografia e Estatística. *Pesquisa Nacional por Amostra de Domicílios: Tabagismo*. Rio de Janeiro: IBGE, 2009. 1–128 p. ISBN 9788524041020.

_____. *Pesquisa de Orçamentos Familiares 2008-2009: despesas, rendimento e condições de vida*. Rio de Janeiro: IBGE, 2010. 1–222 p. ISBN 9788524041310.

_____. *Pesquisa Nacional de Saúde do Escolar*. Rio de Janeiro: IBGE, 2013. 1–256 p. ISBN 9788524042898.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. *Dicionário de indicadores educacionais: fórmulas de cálculo*. Brasília-DF: MEC/INEP, 2004. 1–28 p.

JAMISON, D. T. Child Malnutrition and School Performance in China. *Journal of Development Economics*, v. 20, p. 299–309, 1986.

JEYNES, W. H. The relationship between the consumption of various drugs by adolescents and their academic achievement. *The American Journal of Drug and Alcohol Abuse*, v. 28, n. 1, p. 15–35, jan. 2002. ISSN 0095-2990. Disponível em: <http://informahealthcare.com/doi/abs/10.1081/ADA-120001279>.

JIMENEZ, S.; LABEAGA, J. M. Is it possible to reduce tobacco consumption via alcohol taxation? *Health Economics*, v. 3, n. 4, p. 231–241, 1994.

JOHNSON, J. A.; OKSANEN, E. H. Estimation of Demand for Alcoholic Beverages in Canada From Pooled Time Series and Cross Sections. *The Review of Economics and Statistics*, v. 59, n. 1, p. 113–118, 1977.

JURGES, H.; REINHOLD, S.; SALM, M. Does schooling affect health behavior? Evidence from the educational expansion in Western Germany. *Economics of Education Review*, v. 30, n. 5, p. 862–872, out. 2011. ISSN 02727757.

KAESTNER, R.; GROSSMAN, M. Effects of weight on children's educational achievement. *Economics of Education Review*, v. 28, n. 6, p. 651–661, dez. 2009. ISSN 02727757. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0272775709000557>.

KAESTNER, R.; GROSSMAN, M.; YARNOFF, B. Effects of Weight on Adolescent Educational Attainment. In: GROSSMAN, M.; MOCAN, N. H. (Ed.). *Economic Aspects of Obesity*. Chicago: University of Chicago Press, 2011. p. 283–313. ISBN 9780226310091.

KASSOUF, A. L. Saúde e mercado de trabalho. *Pesquisa e Planejamento Econômico*, v. 27, n. 3, p. 587–610, 1997.

KEBEDE, B. Intra-Household allocations in Rural Ethiopia: a demand systems approach. *Review of Income and Wealth*, v. 54, n. 1, p. 1–26, 2008.

KENKEL, D. S.; WANG, P. Are Alcoholics in Bad Jobs? In: CHALOUPEK, F. J. (Ed.). *The Economic Analysis of Substance Use and Abuse: An Integration of Econometrics and Behavioral Economic Research*. Chicago: University of Chicago Press, 1999.

KOCH, S. F.; RIBAR, D. C. A siblings analysis of the effects of Alcohol consumption onset on Educational Attainment. *Contemporary Economic Policy*, v. 19, n. 2, p. 162–174, 2001.

KOENKER, R. *Quantile Regression*. Cambridge: Cambridge University Press, 2005.

KOENKER, R.; BASSETT, G. Regression Quantiles Roger. *Econometrica*, v. 46, n. 1, p. 33–50, 1978.

KOKSAL, A.; WOHLGENANT, M. Interdependence of Tobacco and Alcohol Consumption: A Natural Experiment Approach. In: *Agricultural & Applied Economics Association's 2013*. Washington, DC: AAEA, 2013. p. 1–26.

KUCZMARSKI, R. J.; OGDEN, C. L.; GUO, S. S.; GRUMMER-STRAWN, L. M.; FLEGAL, K. M.; MEI, Z.; WEI, R.; CURTIN, L. R.; ROCHE, A. F.; JOHNSON, C. L. 2000 CDC Growth Charts for the United States: methods and development. *Vital and health statistics*, v. 11, n. 246, p. 1–190, maio 2002. ISSN 0083-1980.

LEIGH, J. P.; BERGER, M. C. Effects of smoking and being overweight on current earnings. *American Journal of Preventive Medicine*, v. 5, n. 1, p. 8–14, 1989.

LEVINE, P. B.; GUSTAFSON, T. A.; VELENCHIK, A. D. More bad news for smokers? The effects of cigarette smoking on wages. *Industrial and Labor Relations Review*, v. 50, n. 3, p. 493–509, 1997.

LINDEN, W. J. van der; HAMBLETON, R. K. (Ed.). *Handbook of Modern Item Response Theory*. 2. ed. New York: Springer, 2010. 1–510 p.

LINDO, J. M.; SWENSEN, I. D.; WADDELL, G. R. Alcohol and student performance: estimating the effect of legal access. *Journal of health economics*, Elsevier B.V., v. 32, n. 1, p. 22–32, jan. 2013. ISSN 1879-1646. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/23202254>>.

LOKSHIN, M.; BEEGLE, K. Forgone earnings from smoking: evidence for a developing country. *World Bank Policy Research Working Paper n. 4018*, p. 1–27, 2006.

LYE, J.; HIRSCHBERG, J. Alcohol Consumption and Human Capital: a Retrospective Study of the Literature. *Journal of Economic Surveys*, v. 24, n. 2, p. 309–338, abr. 2010. ISSN 09500804. Disponível em: <<http://doi.wiley.com/10.1111/j.1467-6419.2009.00616.x>>.

MACHADO, A.; MORO, S.; MARTINS, L.; RIOS, J. Qualidade do ensino em matemática: determinantes do desempenho de alunos em escolas públicas estaduais mineiras. *Revista da Anpec*, v. 9, n. 1, p. 23–45, 2008. Disponível em: <https://anpec.org.br/revista/vol9/vol9n1p23_45.pdf>.

MACHADO, D. C. Efeitos da saúde na idade de entrada à escola. *Pesquisa e Planejamento Econômico*, v. 38, n. 1, p. 67–96, 2008.

MACHADO, D. C.; GONZAGA, G. O Impacto dos Fatores Familiares Sobre a Defasagem Idade-Série de Crianças no Brasil. *Revista Brasileira de Economia*, v. 61, n. 4, p. 449–476, 2007.

MARTINS, S. S.; ALEXANDRE, P. K. The association of ecstasy use and academic achievement among adolescents in two U.S. national surveys. *Addictive Behaviors*, Elsevier Ltd, v. 34, n. 1, p. 9–16, jan. 2009. ISSN 1873-6327.

MENEZES, T. A.; SILVEIRA, F. G.; AZZONI, C. R. Demand elasticities for food products in Brazil: a two-stage budgeting system. *Applied Economics*, v. 40, n. 19, p. 2557–2572, 2008.

MINCER, J. *Schooling, Experience, and Earnings*. Massachusetts: National Bureau of Economic Research, 1974. 1–152 p. ISBN 0870142658. Disponível em: <<http://www.nber.org/books/minc74-1>>.

- MOOCK, P. R.; LESLIE, J. Childhood malnutrition and schooling in the Terai Region of Nepal. *Journal of Development Economics*, v. 20, p. 33–52, 1986.
- MUELLBAUER, J. Community Preferences and the Representative Consumer. *Econometrica*, v. 44, n. 5, p. 979–999, 1976.
- MUNASINGHE, L.; SICHERMAN, N. Why Do Dancers Smoke? Smoking, Time Preference, and Wage Dynamics. *Eastern Economic Journal*, v. 32, n. 4, p. 595–616, 2006.
- OLIVER, A. S. Information technology and transportation: substitutes or complements? *MPRA paper n. 52896*, p. 1–28, 2014.
- PARK, C.; KANG, C. Does education induce healthy lifestyle? *Journal of Health Economics*, v. 27, n. 6, p. 1516–31, dez. 2008. ISSN 0167-6296.
- PESSOA, D. G. C.; SILVA, P. L. N.; DUARTE, R. P. N. Análise estatística de dados de pesquisas por amostragem: problemas no uso de pacotes-padrões. *Revista Brasileira de Estatística*, v. 58, n. 210, p. 53–75, 1997.
- PINTOS-PAYERAS, J. A. Estimação do Sistema Quase Ideal de Demanda para uma cesta ampliada de produtos empregando dados da POF de 2002-2003. *Economia Aplicada*, v. 13, n. 2, p. 231–255, 2009.
- PIRIE, P. L.; MURRAY, D. M.; LUEPKER, R. V. Smoking Prevalence in a Cohort of Adolescents, Including Absentees, Dropouts, and Transfers. *American Journal of Public Health*, v. 78, n. 2, p. 176–179, 1988.
- POI, B. P. From the help desk : Demand system estimation. *The Stata Journal*, v. 2, n. 4, p. 2002, 2002.
- _____. Demand-system estimation: Update. *The Stata Journal*, v. 8, n. 4, p. 2008, 2008.
- _____. Easy demand-system estimation with quads. *The Stata Journal*, v. 12, n. 3, p. 433–446, 2012.
- POLAT, S. Wage Compensation for Risk: the case of Turkey. *GIAM Working Paper No. 13-11*, n. 13, p. 1–16, 2013.
- POLLAK, R. A.; WALES, T. J. Demographic Variables in Demand Analysis. *Econometrica*, v. 49, n. 6, p. 1533–1551, 1981.
- PONZO, M. Does bullying reduce educational achievement? An evaluation using matching estimators. *Journal of Policy Modeling*, v. 35, n. 6, p. 1057–1078, nov. 2013. ISSN 01618938.
- PSACHAROPOULOS, G.; PATRINOS, H. A. Human capital and rates of return. In: JOHNES, G.; JOHNES, J. (Ed.). *International Handbook on the Economics of Education*. Cheltenham: Edward Elgar Publishing Ltd, 2004. p. 1–57.
- RANGVID, B. S. School composition effects in Denmark: quantile regression evidence from PISA 2000. *Empirical Economics*, v. 33, n. 2, p. 359–388, mar. 2007. ISSN 0377-7332.
- RAY, R. Measuring the costs of children: An alternative approach. *Journal of Public Economics*, v. 22, n. 1, p. 89–102, 1983.

REES, D. I.; SABIA, J. J. The Effect of Migraine Headache on Educational Attainment. *Journal of Human Resources*, v. 46, n. 2, p. 317–332, 2011. ISSN 1548-8004. Disponível em: <http://muse.jhu.edu/content/crossref/journals/journal_of_human_resources/v046/46.2.rees.html>.

_____. The kid's speech: The effect of stuttering on human capital acquisition. *Economics of Education Review*, Elsevier Ltd, v. 38, p. 76–88, fev. 2014. ISSN 02727757. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S027277571300099X>>.

RENNA, F. The economic cost of teen drinking: late graduation and lowered earnings. *Health Economics*, v. 419, p. 407–419, 2007.

_____. Teens' alcohol consumption and schooling. *Economics of Education Review*, v. 27, n. 1, p. 69–78, fev. 2008. ISSN 02727757. Disponível em: <<http://linkinghub.elsevier.com/retrieve/pii/S0272775706001051>>.

RIANI, J. d. L. R.; RIOS-NETO, E. L. G. Background familiar versus perfil escolar do município: qual possui maior impacto no resultado educacional dos alunos brasileiros. *Revista Brasileira de Estudos Populacionais*, v. 25, n. 2, p. 251–269, 2008.

RIBEIRO, R.; CACCIAMALI, M. C. Impactos do Programa Bolsa-Família Sobre os Indicadores Educacionais. *Economia*, v. 13, n. 2, p. 415–446, 2012.

RIOS-NETO, E. L. G.; CESAR, C. C.; RIANI, J. d. L. R. Estratificação educacional e progressão escolar por série no Brasil. *Pesquisa e Planejamento Econômico*, v. 32, n. 3, p. 395–416, 2002.

ROSEN, S. Hedonic Prices and Implicit Markets : Product Differentiation in Pure Competition. *Journal of Political Economy*, v. 82, n. 1, p. 34–55, 1974.

ROSENBAUM, P. R. Attributing to Treatment in Matched Observational Studies. *Journal of the American Statistical Association*, v. 97, n. 457, p. 183–192, 2002.

_____. *Design of Observational Studies*. New York: Springer, 2010. 1–382 p. ISBN 9781441912121.

ROSENBAUM, P. R.; RUBIN, D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, v. 70, n. 1, p. 41–55, 1983.

SAFFER, H.; CHALOUPKA, F. Alcohol tax equalization and social costs. *Eastern Economic Journal*, v. 20, n. 1, p. 33–43, 1994.

SAMPAIO, B.; SAMPAIO, Y.; MELLO, E. P. G. de; MELO, A. S. Desempenho no vestibular, background familiar e evasão: evidências da UFPE. *Economia Aplicada*, v. 15, n. 2, p. 287–309, 2011.

SCHARFF, R. L.; VISCUSI, W. K. Heterogeneous rates of time preference and the decision to smoke. *Economic Inquiry*, v. 49, n. 4, p. 959–972, 2011.

SCHLINDWEIN, M. M.; KASSOUF, A. L. Análise da influência de alguns fatores socioeconômicos e demográficos no consumo domiciliar de carnes no Brasil. *Revista de Economia e Sociologia Rural*, v. 44, n. 3, p. 549–572, set. 2006. ISSN 0103-2003.

- SCHULTZ, T. W. Investment in Human Capital. *The American Economic Review*, v. 51, n. 1, p. 1–17, 1961.
- SCHWARZE, J.; HEINECK, G. Substance use and earnings: the case of smokers in Germany. *IZA Working Paper n. 743*, 2003.
- SHONKWILER, J. S.; YEN, S. T. Two-Step Estimation of a Censored System of Equations. *American Journal of Agricultural Economics*, v. 81, n. 4, p. 972–982, 1999.
- SIGFUSDOTTIR, I. D.; KRISTJANSSON, A. L.; ALLEGRANTE, J. P. Health behaviour and academic achievement in Icelandic school children. *Health Education Research*, v. 22, n. 1, p. 70–80, fev. 2007. ISSN 0268-1153.
- SILVA, P. L. d. N.; PESSOA, D. G. a. C.; LILA, M. F. Análise estatística de dados da PNAD : incorporando a estrutura do plano amostral. *Ciência & Saúde Coletiva*, v. 7, n. 4, p. 659–670, 2002.
- SMALL, K. A.; ROSEN, H. S. Applied Welfare Economics with Discrete Choice Models. *Econometrica*, v. 49, n. 1, p. 105–130, 1981.
- SMITH, R. S. Compensating wage differentials and public policy: a review. *Industrial and Labor Relations Review*, v. 32, n. 3, p. 339–352, 1979.
- SOARES, J. F.; ANDRADE, R. J. de. Nível socioeconômico, qualidade e equidade das escolas de Belo Horizonte. *Avaliação e Políticas Públicas em Educação*, v. 14, n. 40, p. 107–126, 2006.
- SOARES-NETO, J. J.; JESUS, G. R. de; KARINO, C. A.; ANDRADE, D. F. de. Uma escala para medir a infraestrutura escolar. *Estudos em Avaliação Educacional*, v. 24, n. 54, p. 78–99, 2013.
- SOARES, T. M. Influência do professor e do ambiente em sala de aula sobre a proficiência alcançada pelos alunos avaliados no Simave-2002. *Estudos em Avaliação Educacional*, n. 28, p. 103–124, 2003.
- STOCK, J. H.; WRIGHT, J. H.; YOGO, M. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business & Economic Statistics*, v. 20, n. 4, p. 518–529, out. 2002. ISSN 0735-0015. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1198/073500102288618658>>.
- STOCK, J. H.; YOGO, M. Testing for Weak Instruments in Linear IV Regression. In: ANDREWS, D. W. K.; STOCK, J. H. (Ed.). *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press, 2005. p. 80–108.
- STONE, R. Linear Expenditure Systems and Demand Analysis : An Application to the Pattern of British Demand. *The Economic Journal*, v. 64, n. 255, p. 511–527, 1954.
- STRAUSS, J.; THOMAS, D. Health, Nutrition, and Economic Development. *Journal of Economic Literature*, v. 36, n. 2, p. 766–817, 1998.

TEFERA, N.; DEMEKE, M.; RASHID, S. Welfare Impacts of Rising Food Prices in Rural Ethiopia: a Quadratic Almost Ideal Demand System Approach. In: *Proceedings of the International Association of Agricultural Economists Conference*. Foz do Iguaçu: IAAE Triennial Conference, 2012. p. 1–48.

THOMAS, D.; STRAUSS, J. Health and wages: Evidence on men and women in urban Brazil. *Journal of Econometrics*, v. 77, p. 159–185, 1997.

TIMMINS, C.; MURDOCK, J. A revealed preference approach to the measurement of congestion in travel cost models. *Journal of Environmental Economics and Management*, v. 53, n. 2, p. 230–249, mar. 2007. ISSN 00950696. Disponível em: <http://linkinghub.elsevier.com/retrieve/pii/S0095069606000817>.

Van Ours, J. C. A pint a day raises a man's pay; but smoking blows that gain away. *Journal of Health Economics*, v. 23, n. 5, p. 863–86, set. 2004. ISSN 0167-6296. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/15353183>.

VERONNEAU, M.-H.; TREMPE, S.-C.; PAIVA, A. O. Risk and protection factors in the peer context: how do other children contribute to the psychosocial adjustment of the adolescent? *Ciência & Saúde Coletiva*, v. 19, n. 3, p. 695–705, mar. 2014. ISSN 1413-8123.

VISCUSI, W. K.; ALDY, J. E. The Value of a Statistical Life: A Critical Review of Market Estimates Throughout the World. *The Journal of Risk and Uncertainty*, v. 27, n. 1, p. 5–76, 2003.

VISCUSI, W. K.; HERSCH, J. Cigarette smokers as job risk takers. *The Review of Economics and Statistics*, v. 83, n. 2, p. 269–280, 2001.

VISCUSI, W. K.; MOORE, M. J. Workers' Compensation: Wage Effects, Benefit Inadequacies, and the Value of Health Losses. *The Review of Economics and Statistics*, v. 69, n. 2, p. 249–261, 1987.

VU, L.; GLEWWE, P. Impacts of Rising Food Prices on Poverty and Welfare in Vietnam. *Journal of Agricultural and Resource Economics*, v. 36, n. 1, p. 14–27, 2011.

WANG, J.; GAO, X. M.; WAILES, E. J.; CRAMER, G. L. U. S. Consumer Demand for Alcoholic Beverages : Cross-Section Estimation of Demographic and Economic Effects. *Review of Agricultural Economics*, v. 18, n. 3, p. 477–489, 1996.

WHO - World Health Organization. *A Conceptual Framework for Action on the Social Determinants of Health*. Geneva: WHO, 2010. 1–76 p.

_____. Burden: mortality, morbidity and risk factors. In: *Global status report on noncommunicable diseases 2010*. Geneva: WHO, 2011. p. 9–32.

WILLIS, R. J. Wage determinants: a survey and reinterpretation of human capital earnings functions. In: ASHENFELTER, O.; LAYARD, R. (Ed.). *Handbook of Labor Economics*. v. 1. Amsterdam: North Holland, 1986. L, cap. 10, p. 525–602.

WOOD, B. D.; NELSON, C. H.; NOGUEIRA, L. Poverty effects of food price escalation: The importance of substitution effects in Mexican households. *Food Policy*, v. 37, n. 1, p. 77–85, fev. 2012. ISSN 03069192.

WOOLDRIDGE, J. W. *Econometric Analysis of Cross Section and Panel Data*. 2. ed. Massachusetts: MIT Press, 2010. 1–1064 p.

World Health Organization - WHO. *Global health risks: Mortality attributable to selected major risks*. Geneva: Health Statistics and Informatics Department, 2009.

YAMADA, T.; KENDIX, M.; YAMADA, T. The impact of alcohol consumption and marijuana use on high school graduation. *Health Economics*, v. 5, n. 1, p. 77–92, 1996.

YAMAMOTO, C. H. Mestre em Economia. *A demanda por bebidas alcoólicas no Brasil 2008-2009*. Dissertação (Mestrado em Economia): Fundação Getúlio Vargas, 2011. 1–88 p.

YEN, S. T. A Multivariate Sample-Selection Model: Estimating Cigarette and Alcohol Demands with Zero Observations. *American Journal of Agricultural Economics*, v. 87, n. 2, p. 453–466, 2005.

YEN, S. T.; LIN, B.-h.; SMALLWOOD, D. M. Quasi- and Simulated-Likelihood Approaches to Censored Demand Systems: Food Consumption by Food Stamp Recipients in the United States. *American Journal of Agricultural Economics*, v. 85, n. 2, p. 458–478, 2003.

YEN, S. T.; YUAN, Y.; LIU, X. Alcohol consumption by men in China: A non-Gaussian censored system approach. *China Economic Review*, Elsevier Inc., v. 20, n. 2, p. 162–173, jun. 2009. ISSN 1043951X.

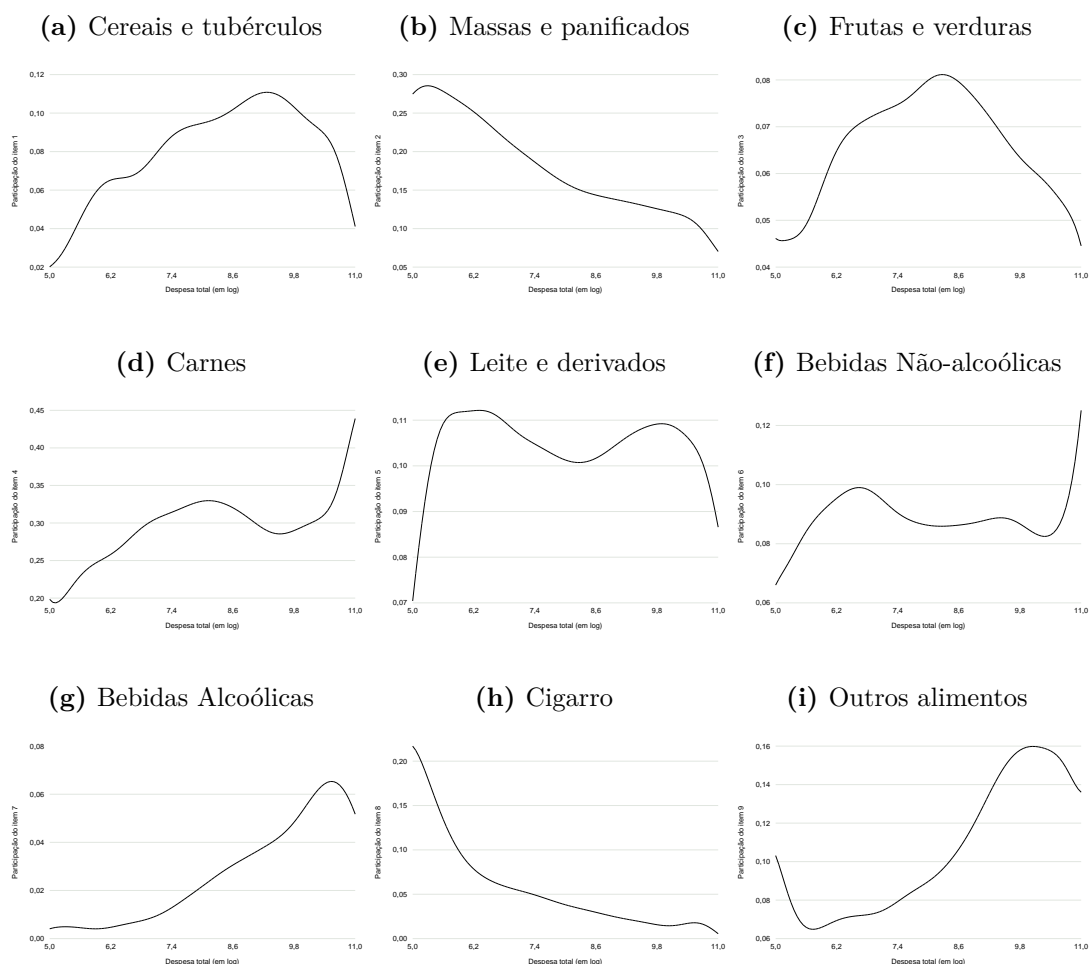
ZARKIN, G. A.; FRENCH, M. T.; MROZ, T.; BRAY, J. W. Alcohol use and wages: new results from the National Household Survey on Drug Abuse. *Journal of Health Economics*, v. 17, p. 53–68, 1998.

ZHENG, Z.; HENNEBERRY, S. R. An Analysis of Food Grain Consumption in Urban Jiangsu Province of China. *Journal of Agricultural and Applied Economics*, v. 42, n. 2, p. 337–355, 2010.

Apêndice

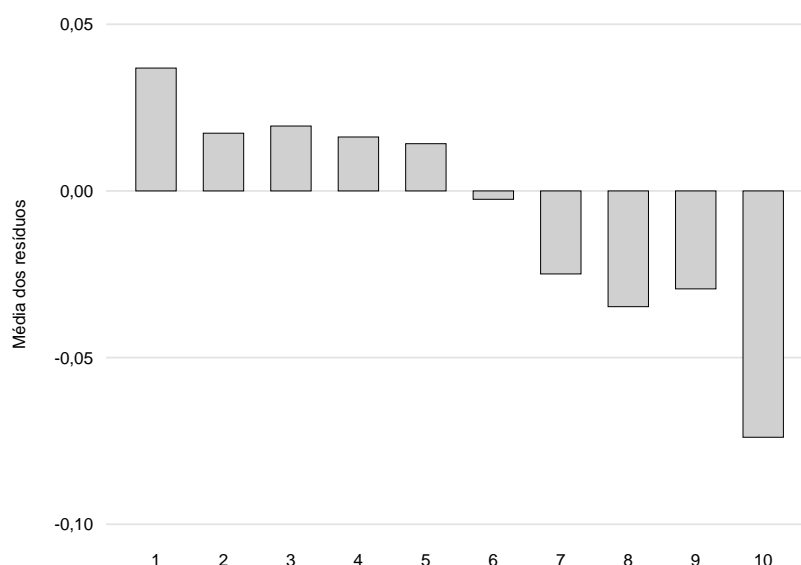
A Apêndice – Primeiro Ensaio

Figura A.1 – Curvas não-paramétricas de Engel por grupos de produtos alimentares, bebidas e cigarro no Brasil 2008-2009



Fonte: Elaboração própria a partir dos microdados da POF 2008-2009.

Nota: Regressão não-paramétrica usando a abordagem *Linear Locally Estimation* com largura de banda de 0,40. Variável dependente é a participação do item nos gastos totais.

Figura A.2 – Resíduos da equação reduzida das despesas totais por decil de renda**Tabela A.1** – Estimativas do modelo probit multivariado – Estágio de correção dos valores censurados

Grupos→ Variáveis↓	1	2	3	4	5	6	7	8	9
lnm	0,8266 ^c	0,4819 ^c	0,5936 ^c	0,7629 ^c	0,6745 ^c	0,6303 ^c	0,6291 ^c	0,1582 ^c	0,8452 ^c
lnp1	-0,6021 ^c	0,1273 ^b	-0,4776 ^c	-0,1128 ^b	0,0192	0,1074 ^c	-0,043	0,0661 ^a	0,0556
lnp2	-0,2780 ^c	-0,4330 ^c	0,1500 ^c	-0,0939 ^b	0,2540 ^c	0,1247 ^c	-0,0028	-0,0600 ^b	-0,0227
lnp3	-0,1894 ^c	-0,0241	-0,5664 ^c	0,0858 ^b	-0,0197	0,0583 ^b	0,0749 ^c	0,0925 ^c	-0,0059
lnp4	-0,2448 ^c	0,0001	-0,1734 ^c	-0,4319 ^c	-0,0377	-0,1403 ^c	0,0194	-0,0689 ^b	-0,1756 ^c
lnp5	0,0216	-0,0146	-0,004	0,031	-0,4953 ^c	0,0727 ^c	0,0195	0,0224	0,1029 ^c
lnp6	0,1245 ^c	-0,0564 ^c	-0,0087	-0,0547 ^c	0,0099	-0,4186 ^c	-0,0732 ^c	0,0395 ^c	0,0827 ^c
lnp7	0,0937	0,0768	-0,0403	-0,0803	0,0462	-0,1098	-0,8660 ^c	-0,1405 ^b	-0,0472
lnp8	-0,039	0,2362 ^c	0,2437 ^c	-0,1377 ^a	0,0751	-0,0888	0,1124 ^a	-0,3759 ^c	0,0179
lnp9	-0,3112 ^c	-0,0075	0,1246 ^c	0,0062	-0,0132	-0,1767 ^c	0,0936 ^c	0,0381 ^b	-0,4547 ^c
esc _C	-0,0077 ^b	0,0071	0,0088 ^b	-0,0230 ^c	0,0316 ^c	0,0003	-0,0045	-0,0352 ^c	0,0070 ^a
idade _C	-0,0033	-0,0176 ^c	0,007	-0,0115 ^b	-0,0197 ^c	-0,0112 ^b	0,0039	0,0368 ^c	-0,0180 ^c
idade _C ²	0,0027	0,0173 ^c	-0,0013	0,007	0,0203 ^c	0,0059	-0,0102 ^b	-0,0395 ^c	0,0139 ^c
mulher _C	0,0677 ^c	0,0572	0,1139 ^c	-0,0237	0,1385 ^c	0,005	-0,1953 ^c	0,0498 ^b	0,0489 ^a
ocupacao _C	0,0292	-0,0235	0,0796 ^b	-0,0273	0,0309	-0,0770 ^a	-0,1378 ^c	-0,0632 ^a	-0,0357
analf _D	-0,0505	-0,3684 ^c	-0,0668	0,1819 ^b	-0,2470 ^c	-0,1495 ^b	-0,0602	0,1044 ^a	0,0937
criancas _{1D}	-0,0399 ^c	0,0703 ^c	-0,0478 ^c	-0,0219	0,0244 ^a	-0,0351 ^b	-0,021	0,0608 ^c	-0,0076
criancas _{2D}	-0,0633 ^c	0,0278	-0,0819 ^c	0,0195	-0,0749 ^c	0,0624 ^c	-0,0461 ^b	0,0310 ^a	-0,0173
ocupacao _D	-0,1725 ^c	-0,0337	-0,1249 ^b	-0,1197 ^a	-0,0573	0,2715 ^c	0,5325 ^c	0,3110 ^c	0,0838
urbana	0,0487 ^a	0,4910 ^c	0,1119 ^c	0,0975 ^c	0,2968 ^c	0,0683 ^b	-0,0858 ^c	0,0346	0,0221
reg1	0,1329 ^c	-0,2017 ^c	0,2766 ^c	0,2915 ^c	0,1070 ^b	0,0194	-0,1934 ^c	-0,2605 ^c	-0,0179
reg2	0,2244 ^c	0,1276 ^c	0,2682 ^c	0,2490 ^c	0,2263 ^c	-0,0614 ^a	0,0287	-0,2446 ^c	0,0125
reg4	0,0108	-0,1625 ^c	0,1407 ^c	0,011	0,1567 ^c	0,0344	-0,0196	0,0055	0,2403 ^c
reg5	0,0852 ^b	-0,3754 ^c	0,0777 ^b	0,0862 ^b	-0,1963 ^c	-0,0756 ^b	-0,1117 ^c	-0,1170 ^c	-0,0600 ^a
carnaval	-0,017	-0,0193	-0,0552	0,1017 ^a	-0,0397	-0,0728	-0,0192	0,0065	-0,0459
juninas	0,1141 ^c	0,0574	0,0104	0,0538	-0,0124	-0,0367	0,0631	0,0562	0,0378
natal	-0,0472	-0,0248	-0,0233	0,0079	0,0185	-0,1157 ^b	0,0891 ^a	-0,0958 ^b	-0,1381 ^c
constante	-4,5388 ^c	-1,1171 ^c	-3,4292 ^c	-3,4165 ^c	-4,1349 ^c	-3,1377 ^c	-4,5406 ^c	-3,2708 ^c	-4,7036 ^c

Fonte: Elaboração própria a partir dos microdados da POF 2008-2009.

Nota: Estimativas ajustadas pelo o fator de expansão amostral da POF. Para o teste de razão de verossimilhança de igualdade de $r_{21}, r_{31}, \dots, r_{98}$ à zero, verifica-se que $\chi^2_{(36)} = 2,9e + 08$, com $Pr > \chi^2 = 0,0000$.

Os sobrescritos *a*, *b* e *c* significam, respectivamente: p-valor<0,1; p-valor<0,05; p-valor<0,01.

Tabela A.2 – Matriz de correlação dos termos de erros das equações no modelo probit multivariado – Estágio de correção dos valores censurados

Equações	1	2	3	4	5	6	7	8
2	0,0204							
3	0,1990***	0,0136						
4	-0,0343*	-0,0565**	0,0753***					
5	-0,0091	0,2006***	0,0514***	-0,0537***				
6	-0,0118	0,0258	-0,0136	-0,1169***	0,0702***			
7	-0,1479***	-0,1449***	-0,1562***	-0,1790***	-0,1756***	-0,0582***		
8	-0,1144***	-0,0665***	-0,0440***	-0,0901***	-0,0832***	0,0474***	0,1601***	
9	0,1831***	0,0901***	0,0631***	-0,1297***	0,1190***	0,1746***	-0,1456***	-0,0617***

Fonte: Elaboração própria a partir dos microdados da POF 2008-2009. Estimativas ajustadas pelo o fator de expansão amostral da POF.

Tabela A.3 – Parâmetros do QUAIDS sem ajustes para consumo censurado e endogeneidade das despesas totais

Parâmetros	Equações das composições dos grupos de produtos							
	Cereais(1)	Massas(2)	Frutas(3)	Carnes(4)	Leite(5)	Não-Alcool(6)	Álcool(7)	Cigarro(8)
α_i	0,051788***	0,363972***	0,047573***	0,133767***	0,126165***	0,134683***	0,012893***	0,081495***
β_i	0,041877***	-0,106546***	0,021210***	0,082022***	-0,021513***	-0,011224***	0,014848***	-0,047207***
γ_{i1}	0,014366***	-0,002284**	-0,006592***	-0,008938***	0,002733***	0,015342***	0,006271***	0,002790**
γ_{i2}		-0,017900***	0,015234***	-0,000355*	0,003934***	-0,001951***	0,008263***	0,000128
γ_{i3}			-0,032792***	-0,006389***	-0,000389	0,003569***	0,004812***	0,011648***
γ_{i4}				0,012616	0,003004***	0,003953***	0,002569***	-0,003029*
γ_{i5}					-0,018854***	0,001913***	0,002431***	-0,003504***
γ_{i6}						-0,024293***	-0,000734**	0,000486
γ_{i8}							-0,031508***	0,003647***
γ_{i7}								-0,013997***
λ_i								0,004231***
$\eta_i(\text{esc}_C)$	-0,000143	0,011232***	-0,005932***	-0,018486***	0,003512***	0,001222***	0,000963***	0,000340***
$\eta_i(\text{idade}_C)$	-0,000742***	0,000131**	0,000384***	-0,000983***	0,000973***	0,000149***	0,000062*	0,000608***
$\eta_i(\text{idade}_{2C})$	-0,000485***	0,000078	0,000180***	0,000664***	-0,000336***	-0,000023	-0,000111***	0,000613***
$\eta_i(\text{idade}_{2D})$	0,000342***	0,000023	-0,000003	-0,000595***	0,000410***	0,000015	0,000004	-0,000613***
$\eta_i(\text{mulher}_C)$	-0,000545	0,002464***	0,000575**	-0,001198*	0,001243***	0,000216	-0,002284***	0,000679*
$\eta_i(\text{analf}_D)$	0,000183	-0,001009	-0,002475**	0,004935***	-0,001584*	-0,000464	-0,001907***	0,001026
$\eta_i(\text{criancas}_{1D})$	0,000120	0,003078***	-0,000867***	-0,001080***	0,001876***	-0,001166***	-0,000761***	-0,000759***
$\eta_i(\text{criancas}_{2D})$	0,001251***	0,002592***	-0,001509***	0,000106	-0,002207***	0,000484**	-0,000453***	-0,001119***
$\eta_i(\text{urbana})$	-0,001403***	0,001605***	0,000063	-0,002800***	0,002488***	-0,001386***	0,000676**	0,000739*
$\eta_i(\text{reg1})$	-0,006644***	0,002076***	0,001740***	0,027885***	-0,006272***	0,001628***	-0,002828***	-0,004654***
$\eta_i(\text{reg2})$	-0,002881***	0,007645***	0,001875**	0,018297***	-0,002486***	-0,005523***	-0,000936***	-0,001119***
$\eta_i(\text{reg4})$	-0,005172***	0,003191***	-0,000381	0,000779	-0,001161**	0,002963***	-0,000406	0,000962
$\eta_i(\text{reg5})$	0,002819***	-0,003595***	0,003290***	0,011283***	-0,006256***	0,001575***	-0,002121***	-0,002023***
$\rho_i(\text{esc}_C)$	0,010359***	0,010359***	0,010359***	0,010359***	0,010359***	0,010359***	0,010359***	0,010359***
$\rho_i(\text{idade}_C)$	-0,013990***	-0,013990***	-0,013990***	-0,013990***	-0,013990***	-0,013990***	-0,013990***	-0,013990***
$\rho_i(\text{idade}_{2C})$	0,010914***	0,010914***	0,010914***	0,010914***	0,010914***	0,010914***	0,010914***	0,010914***
$\rho_i(\text{mulher}_C)$	-0,093641***	-0,093641***	-0,093641***	-0,093641***	-0,093641***	-0,093641***	-0,093641***	-0,093641***
$\rho_i(\text{analf}_D)$	-0,081255**	-0,081255**	-0,081255**	-0,081255**	-0,081255**	-0,081255**	-0,081255**	-0,081255**
$\rho_i(\text{criancas}_{1D})$	0,034492***	0,034492***	0,034492***	0,034492***	0,034492***	0,034492***	0,034492***	0,034492***
$\rho_i(\text{criancas}_{2D})$	0,084563***	0,084563***	0,084563***	0,084563***	0,084563***	0,084563***	0,084563***	0,084563***
$\rho_i(\text{urbana})$	0,124612***	0,124612***	0,124612***	0,124612***	0,124612***	0,124612***	0,124612***	0,124612***
$\rho_i(\text{reg1})$	-0,296037***	-0,296037***	-0,296037***	-0,296037***	-0,296037***	-0,296037***	-0,296037***	-0,296037***
$\rho_i(\text{reg2})$	-0,269352***	-0,269352***	-0,269352***	-0,269352***	-0,269352***	-0,269352***	-0,269352***	-0,269352***
$\rho_i(\text{reg4})$	-0,054165	-0,054165	-0,054165	-0,054165	-0,054165	-0,054165	-0,054165	-0,054165
$\rho_i(\text{reg5})$	-0,269314***	-0,269314***	-0,269314***	-0,269314***	-0,269314***	-0,269314***	-0,269314***	-0,269314***
N	37,671	37,671	37,671	37,671	37,671	37,671	37,671	37,671
P	39.939.352	39.939.352	39.939.352	39.939.352	39.939.352	39.939.352	39.939.352	39.939.352

Fonte: Elaboração própria a partir dos microdados da POF 2008-2009 e da PNAD 2008.

Nota: A equação omitida do sistema refere-se ao grupo de outros alimentos, onde os parâmetros são obtidos através das condições de aditividade, simetria de Slutsky e homogeneidade. *p-valor<0,1; **p-valor<0,05; ***p-valor<0,01.

B Apêndice – Segundo Ensaio

Tabela B.1 – Descrição dos itens utilizados para a elaboração do Nível Socioeconômico (NSE)

Item	Descrição do item	Escala	Prop.	a	d
1	Existe telefone fixo (convencional)	1, se sim; 0, c.c.	0,528	1,193	0,017
2	O aluno tem celular	1, se sim; 0, c.c.	0,850	0,742	2,076
3	Possui computador em casa	1, se sim; 0, c.c.	0,660	6,371	3,160
4	Acesso à internet em casa	1, se sim; 0, c.c.	0,604	4,890	1,897
5	Algum morador na residência possui carro	1, se sim; 0, c.c.	0,502	1,613	0,211
6	Algum morador na residência possui moto	1, se sim; 0, c.c.	0,297	-0,112	-0,688
7	<i>Número de banheiros com chuveiro dentro da casa^a</i>	De 0 a 4		1,180	
	Sem banheiro	0	0,043	0,000	0,000
	Apenas um banheiro	1	0,649	1,000	3,436
	Dois banheiros	2	0,228	2,000	2,615
	Três banheiros	3	0,055	3,000	0,615
	Quatro banheiros ou mais	4	0,025	4,000	-1,612
8	Empregado doméstico na casa cinco ou mais dias por semana	1, se sim; 0, c.c.	1,480	1,465	-2,502
9	<i>Faixa de escolaridade da mãe^a</i>	De 0 a 4		0,749	
	Nunca estudou ou ensino fundamental incompleto	0	0,423	0,000	0,000
	Ensino fundamental completo ou médio incompleto	1	0,181	1,000	-0,463
	Ensino médio completo	2	0,234	2,000	-0,174
	Ensino superior incompleto	3	0,054	3,000	-1,818
	Ensino superior completo ou mais anos de estudo	4	0,109	4,000	-1,702
10	<i>Faixa de escolaridade do pai^a</i>	De 0 a 4		0,724	
	Nunca estudou ou ensino fundamental incompleto	0	0,468	0,000	0,000
	Ensino fundamental completo ou médio incompleto	1	0,176	1,000	-0,653
	Ensino médio completo	2	0,207	2,000	-0,470
	Ensino superior incompleto	3	0,045	3,000	-2,151
	Ensino superior completo ou mais anos de estudo	4	0,105	4,000	-1,899

Fonte: Elaboração própria a partir dos microdados da PENSE 2012, tendo por base Soares & Andrade (2006) e Alves & Soares (2012).

Nota: Parâmetros estimados pela TRI policotômica ordinal unidimensional: a =inclinação; d =intercepto. Nesse caso o modelo usado foi o de crédito parcial generalizado, obtido através da seguinte equação:

$$Pr_i q(\theta) = \frac{\exp[\sum_{j=0}^q Ba_i(\theta - d_{ij})]}{\sum_{r=0}^{m_i} [\exp \sum_{j=0}^r Ba_i(\theta - d_{ij})]}, \text{ onde } \theta \text{ é o traço latente e o indexador } i \text{ refere-se ao } i\text{-ésimo item, com } m \text{ categorias.}$$

^a Variável policotômica.

Tabela B.2 – Descrição dos itens utilizados para a elaboração do indicador de infraestrutura da escola (IEE)

Item	Descrição do item	Escala	Prop.	a	d
1	Existência de biblioteca	1, se sim; 0, c.c.	0,908	1,275	2,481
2	Existência de laboratório de informática	1, se sim; 0, c.c.	0,871	1,044	2,188
3	Acesso direto à computadores na sala de aula	1, se sim; 0, c.c.	0,279	0,430	-1,234
4	Acesso à internet através de equipamentos da escola	1, se sim; 0, c.c.	0,859	1,061	1,335
5	Existência de sala para os recursos de mídia	1, se sim; 0, c.c.	0,683	1,008	0,512
6	Existência de cantina	1, se sim; 0, c.c.	0,552	0,959	-0,069
7	Existência de quadra de esportes	1, se sim; 0, c.c.	0,830	5,705	4,590
8	Existência de quadra de esporte coberta	1, se sim; 0, c.c.	0,745	3,309	0,699
9	Existência de pista para corrida/atletismo	1, se sim; 0, c.c.	0,026	1,400	-4,502
10	Existência de piscina em condições de uso	1, se sim; 0, c.c.	0,140	1,943	-3,338
11	Existência de vestiário em condições de uso para os alunos	1, se sim; 0, c.c.	0,358	1,839	-1,387

Fonte: Elaboração própria a partir dos microdados da PENSE 2012, tendo por base Soares-Neto *et al.* (2013).

Nota: Parâmetros estimados pela TRI dicotômica unidimensional: a =inclinação; d =intercepto. O modelo logístico de dois parâmetros para o caso de resposta binária é expresso por:

$Pr_i(\theta) = 1/\{1 + \exp[-a_i(\theta + d_i)]\}$, onde θ é o traço latente e o indexador i refere-se ao i -ésimo item.

Tabela B.3 – Estatística descritivas das variáveis

Variável	N	P	Média	DP	Mín.	Máx.
Gênero (homem=1)	66.844	1.897.670	0,480	0,500	0,000	1,000
Raça (Branca=1)	66.844	1.897.670	0,386	0,487	0,000	1,000
Trabalha	66.844	1.897.670	0,129	0,336	0,000	1,000
Experiência: cigarro	66.844	1.897.670	0,180	0,384	0,000	1,000
Experiência e uso: cigarro	66.844	1.897.670	0,041	0,198	0,000	1,000
Experiência: álcool	66.844	1.897.670	0,663	0,473	0,000	1,000
Experiência e uso: álcool	66.844	1.897.670	0,244	0,429	0,000	1,000
Sobrepeso	66.844	1.897.670	0,158	0,365	0,000	1,000
Núcleo familiar biparental	66.844	1.897.670	0,679	0,467	0,000	1,000
NSE do aluno	66.844	1.897.670	0,458	0,199	0,000	1,000
Tamanho da família	66.844	1.897.670	4,583	1,614	1,000	30,000
Pai fumante	66.844	1.897.670	0,139	0,346	0,000	1,000
Mãe fumante	66.844	1.897.670	0,085	0,279	0,000	1,000
Reação contrária dos pais ao uso do álcool	66.844	1.897.670	0,906	0,292	0,000	1,000
Frequência de consumo: legumes	66.844	1.897.670	3,397	2,590	0,000	7,000
Frequência de consumo: frutas	66.844	1.897.670	3,100	2,495	0,000	7,000
Tamanho da escola (matrículas em mil)	66.844	1.897.670	0,983	0,568	0,067	8,300
Rede pública	66.844	1.897.670	0,802	0,399	0,000	1,000
Conselho Escolar	66.844	1.897.670	0,867	0,340	0,000	1,000
Índice de infraestrutura (IEE)	66.844	1.897.670	0,480	0,196	0,000	1,000
NSE da escola	66.844	1.897.670	0,455	0,145	0,105	0,941
<i>Prevalência de exposição na escola</i>						
Experiência: cigarro (def.1)	66.844	1.897.670	0,194	0,107	0,000	0,714
Experiência e uso: cigarro (def.2)	66.844	1.897.670	0,050	0,049	0,000	0,538
Experiência: álcool (def.1)	66.844	1.897.670	0,669	0,132	0,050	1,000
Experiência e uso: álcool (def.2)	66.844	1.897.670	0,259	0,106	0,000	0,722
Sobrepeso	66.844	1.897.670	0,156	0,068	0,000	0,600
Capital	66.844	1.897.670	0,780	0,415	0,000	1,000
Região Norte	66.844	1.897.670	0,082	0,275	0,000	1,000
Região Nordeste	66.844	1.897.670	0,257	0,437	0,000	1,000
Região Sul	66.844	1.897.670	0,157	0,364	0,000	1,000
Região Centro-Oeste	66.844	1.897.670	0,080	0,271	0,000	1,000

Fonte: Elaboração própria a partir dos microdados da PENSE 2012.

Tabela B.4 – Efeito marginal de variáveis explicativas discretas (D) e contínuas (C) sobre a probabilidade de atraso escolar no Brasil 2012 – a partir do modelo logit do Estágio I

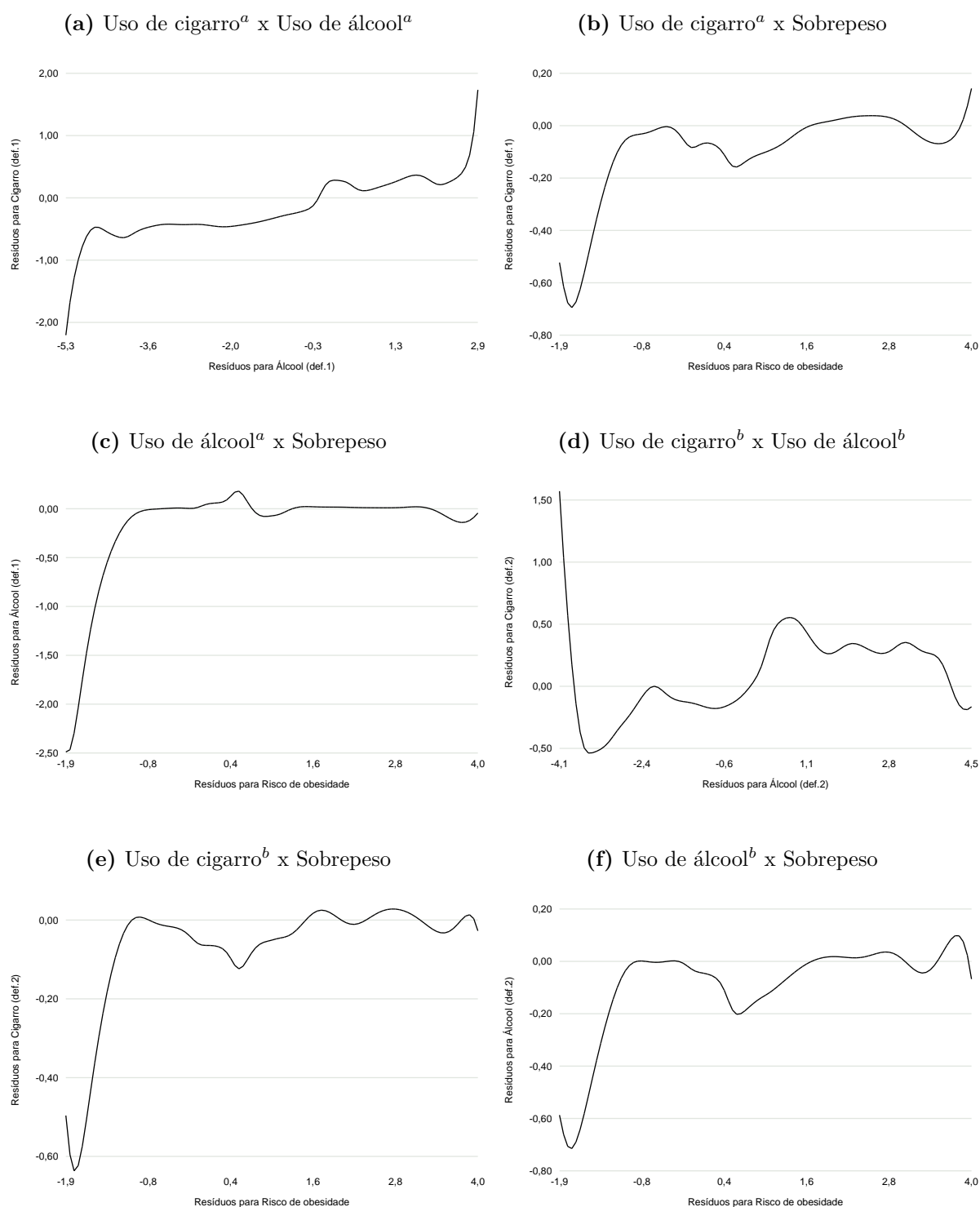
Covariadas	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Atributos do aluno e da família									
D. Gênero (homem=1)	0,0937***	0,0965***	0,0942***	0,0945***	0,0966***	0,0954***	0,0954***	0,0967***	0,0968***
D. Raça (Branca=1)	-0,0276***	-0,0288***	-0,0287***	-0,0280***	-0,0279***	-0,0278***	-0,0278***	-0,0276***	-0,0276***
D. Trabalho	0,0968***	0,1009***	0,1066***	0,1018***	0,0918***	0,0942***	0,0941***	0,0901***	0,0901***
D. Núcleo familiar biparental	-0,0740***	-0,0767***	-0,0809***	-0,0782***	-0,0736***	-0,0721***	-0,0722***	-0,0726***	-0,0726***
C. NSE do aluno	-0,2592***	-0,2697***	-0,2617***	-0,2614***	-0,2689***	-0,2641***	-0,2641***	-0,2678***	-0,2680***
C. Tamanho da família	0,0144***	0,0151***	0,0148***	0,0145***	0,0144***	0,0146***	0,0146***	0,0143***	0,0143***
Atributos da escola									
C. Total de matrículas (em mil)	0,0088	0,0092	0,0097	0,0092	0,0095	0,0084	0,0084	0,0091	0,0091
D. Rede pública	-0,0167	-0,0127	-0,0115	-0,0134	-0,0159	-0,0167	-0,0168	-0,0165	-0,0165
D. Conselho Escolar	-0,0478***	-0,0469***	-0,0468***	-0,0472***	-0,0467***	-0,0477***	-0,0477***	-0,0468***	-0,0469***
C. Índice de infraestrutura (IEE)	-0,1289***	-0,1281***	-0,1276***	-0,1282***	-0,1289***	-0,1287***	-0,1286***	-0,1286***	-0,1287***
C. NSE da escola	-0,3886***	-0,3920***	-0,3804***	-0,3826***	-0,3854***	-0,3952***	-0,3955***	-0,3866***	-0,3863***
Fatores regionais									
D. Capital	-0,0606	-0,0630	-0,0638	-0,0626	-0,0629	-0,0603	-0,0603	-0,0621	-0,0620
D. Região Norte	0,0286	0,0285	0,0263	0,0283	0,0315	0,0304	0,0304	0,0329	0,0329
D. Região Nordeste	-0,0207	-0,0221	-0,0224	-0,0207	-0,0216	-0,0197	-0,0197	-0,0198	-0,0198
D. Região Sul	-0,0697	-0,0647	-0,0602	-0,0634	-0,0676	-0,0711	-0,0710	-0,0684	-0,0685
D. Região Centro-Oeste	0,0060	0,0094	0,0110	0,0095	0,0087	0,0063	0,0064	0,0088	0,0089
Exposição aos fatores de risco									
D. Experiência: cigarro (def.1)	0,0866***				0,0759***	0,0759***	0,0695***		
D. Experiência: álcool (def.1)		0,0452***			0,0286	0,0286	0,0318**		
D. Sobre peso			0,0154**		0,0144*	0,0144*	0,0144*	0,0148**	0,0146**
D. Experiência e uso: cigarro (def.2)								0,0528***	0,0848***
D. Experiência e uso: álcool (def.2)				0,1086***	0,1018***			0,0951***	0,0949***
Observações	66.844	66.844	66.844	66.844	66.844	66.844	66.844	66.844	66.844
População	1.897.670	1.897.670	1.897.670	1.897.670	1.897.670	1.897.670	1.897.670	1.897.670	1.897.670
Estratos	32	32	32	32	32	32	32	32	32
UPA	1.473	1.473	1.473	1.473	1.473	1.473	1.473	1.473	1.473

Fonte: Elaboração própria a partir dos microdados da PENSE 2012.

Nota: As estimativas são desenvolvidas considerando, além do peso amostral, a estratificação e a UPA do plano amostral. D = estimativa calculada pela Equação 3.4; C = estimativa calculada pelo efeito parcial médio (EPM).

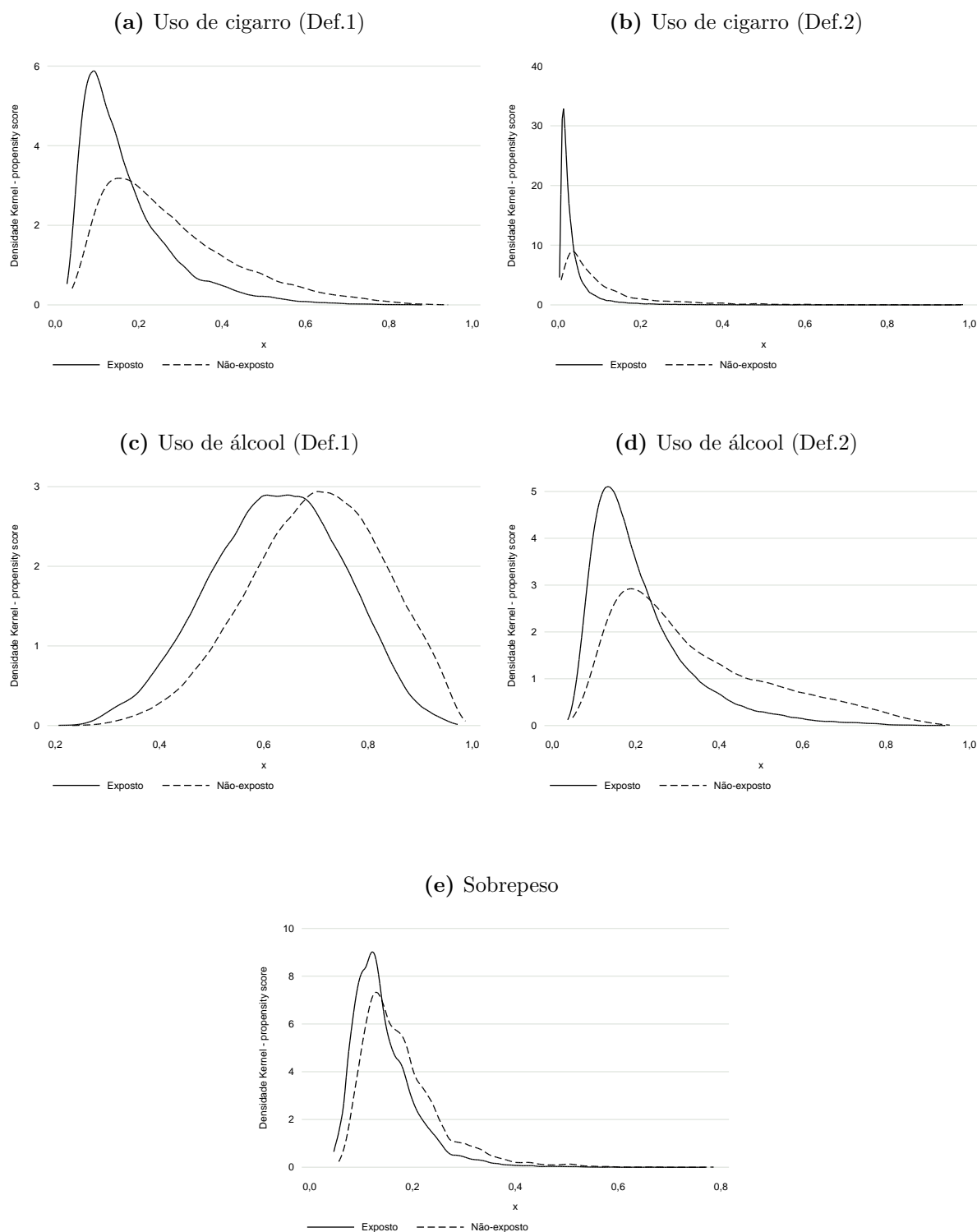
Legenda: *p-valor<0,10; **p-valor<0,05; ***p-valor<0,01.

Figura B.1 – Curvas não-paramétricas entre os termos residuais das equações do estágio inicial do PSM



Fonte: Elaboração própria a partir dos microdados da PENSE 2012. Legenda: a = definição 1; b = definição 2

Figura B.2 – Densidade Kernel do *propensity score* entre expostos e não-expostos por tipo de fator de risco comportamental à saúde



Fonte: Elaboração própria a partir dos microdados da PENSE 2012.

C Apêndice – Terceiro Ensaio

Tabela C.1 – Descrição das estatísticas do total de acidentes do trabalho (AT) no Brasil por atividade econômica e do índice de risco de acidentes do trabalho (IRAT), 2008

	Nome da divisão CNAE-Domiciliar	AT (A)	Ocupações (B)	IRAT* (A/B)*
1	Fabricação de máquinas para escritório e equipamentos de informática	4.087	45.277	9,03
2	Reciclagem	8.665	98.585	8,79
3	Fabricação de coque, refino de petróleo, elaboração de combustíveis nucleares e produção de álcool	8.211	186.625	4,40
4	Fabricação de artigos de borracha e plástico	15.849	360.401	4,40
5	Fabricação e montagem de veículos, reboques e carrocerias	26.339	631.144	4,17
6	Fabricação de celulose, papel e produtos de papel	6.945	207.970	3,34
7	Fabricação de máquinas e equipamentos	18.437	592.133	3,11
8	Fabricação de produtos do fumo	505	16.359	3,09
9	Fabricação de máquinas, aparelhos e materiais elétricos	6.803	222.772	3,05
10	Fabricação de produtos alimentícios e bebidas	69.923	2.329.237	3,00
11	Fabricação de outros equipamentos de transporte	4.394	148.043	2,97
12	Extração de minerais metálicos	2.904	99.623	2,91
13	P&D das ciências sociais e humanas	1.475	51.817	2,85
14	Metalurgia básica	14.980	552.479	2,71
15	Fabricação de produtos de metal - excl. máquinas e equipamentos	20.055	881.636	2,27
16	Transporte aéreo	1.099	49.917	2,20
17	Correios e telecomunicações	14.063	681.962	2,06
18	Fabricação de produtos de minerais não-metálicos	12.717	665.942	1,91
19	Captação, tratamento e distribuição de água	3.224	171.746	1,88
20	Fabricação de produtos de madeira	9.187	492.897	1,86
21	Eletricidade, gás e água quente	3.280	201.525	1,63
22	Saúde e serviços sociais	53.695	3.415.903	1,57
23	Atividades associativas	10.849	719.641	1,51
24	Atividades anexas e auxiliares do transporte e agências de viagem	7.356	492.266	1,49
25	Aluguel de veículos, máquinas e equipamentos sem condutores ou operadores e de objetos pessoais	2.266	152.676	1,48
26	Fabricação de produtos químicos	8.951	645.166	1,39
27	Extração de minerais não-metálicos	2.404	179.576	1,34
28	Transporte aquaviário	706	53.740	1,31
29	Silvicultura, exploração florestal e serviços relacionados	4.330	346.643	1,25
30	Fabricação de móveis e indústrias diversas	10.848	942.548	1,15
31	Preparação de couros e fabricação de artefatos de couro	9.284	825.214	1,13
32	Intermediação financeira, excl. seguros e previdência privada	9.080	809.559	1,12
33	Fabricação de produtos têxteis	9.708	889.388	1,09
34	Serviços prestados principalmente às empresas	45.371	4.387.884	1,03
35	Limpeza urbana e esgoto; e atividades conexas	2.081	209.759	0,99
36	Extração de petróleo, gás natural e serviços relacionados	736	76.193	0,97
37	Transporte terrestre	31.407	3.308.769	0,95
38	Seguros e previdência privada	1.490	172.653	0,86
39	Edição, impressão e reprodução de gravações	3.650	438.189	0,83
40	Construção	52.830	6.896.488	0,77
41	Comércio a varejo e por atacado; e reparação de objetos pessoais	87.226	13.099.192	0,67
42	Confecção de artigos do vestuário e acessórios	12.181	1.886.834	0,65
43	Alojamento e alimentação	18.797	3.579.661	0,53
44	Administração pública, defesa e seguridade social	21.874	4.514.317	0,48
45	Atividades de informática e conexas	2.852	618.412	0,46
46	Comércio e reparação de veículos automotores e motocicletas; e comércio a varejo de combustíveis	11.924	2.947.457	0,40
47	Atividades recreativas, culturais e desportivas	3.165	1.244.329	0,25
48	Atividades auxiliares da intermediação financeira	374	170.398	0,22
49	Educação	8.411	5.104.221	0,16
50	Agricultura, pecuária e serviços relacionados	25.022	15.485.414	0,16
51	Atividades imobiliárias	886	751.272	0,12
52	Serviços pessoais	1.752	1.896.458	0,09
53	Pesca, aquicultura e atividades dos serviços relacionados	358	406.616	0,09
54	Serviços domésticos	25	6.611.902	0,00

Fonte: Elaboração própria a partir dos microdados da PNAD/PETab 2008 e do AEAT/MPS 2008.

Nota: *Índice de risco de acidentes do trabalho (IRAT) por 100 trabalhadores. As atividades com menos de 30 observações (sem expansão amostral) foram desconsideradas do cálculo referente à quantidade de ocupações pela PNAD.

Tabela C.2 – Resultados para a penalização salarial do fumante corrente para especificações com ocupação informal e ramos da atividade econômica – MQO e MQ2E. Variável dependente: logaritmo do salário-hora

Covariadas	MQO		MQ2E	
	(3)	(3')	(6)	(6')
Fumante	-0,1053*** (0,0212)	-0,0944*** (0,0212)	-0,2479*** (0,0526)	-0,2304*** (0,0532)
Faixa de escolaridade				
<EF	-0,3778*** (0,0843)	-0,3396*** (0,0837)	-0,3648*** (0,0844)	-0,3294*** (0,0839)
<EM	-0,3818*** (0,0682)	-0,3516*** (0,0679)	-0,3772*** (0,0686)	-0,3481*** (0,0683)
<ES	-0,3300*** (0,0584)	-0,3099*** (0,0581)	-0,3334*** (0,0591)	-0,3129*** (0,0588)
Raça	0,1020*** (0,0201)	0,0984*** (0,0200)	0,0971*** (0,0201)	0,0942*** (0,0199)
Idade	0,0308*** (0,0096)	0,0296*** (0,0095)	0,0306*** (0,0095)	0,0295*** (0,0095)
Idade ² /100	-0,0401*** (0,0116)	-0,0390*** (0,0115)	-0,0382*** (0,0116)	-0,0372*** (0,0115)
Interação: Educação e Idade	0,0011*** (0,0001)	0,0011*** (0,0001)	0,0011*** (0,0001)	0,0011*** (0,0001)
Sindicalizado	0,1248*** (0,0238)	0,1011*** (0,0243)	0,1156*** (0,0238)	0,0953*** (0,0242)
Estabilidade (tenure)	0,0095*** (0,0014)	0,0092*** (0,0014)	0,0094*** (0,0014)	0,0090*** (0,0014)
Grupos ocupacionais				
Dirigentes	0,7921*** (0,0544)	0,8625*** (0,0565)	0,7881*** (0,0553)	0,8587*** (0,0572)
Ciências e Artes	0,7320*** (0,0809)	0,7533*** (0,0797)	0,7221*** (0,0819)	0,7430*** (0,0806)
Técnicos	0,5782*** (0,0440)	0,6133*** (0,0451)	0,5789*** (0,0442)	0,6135*** (0,0451)
Serviços administrativos	0,1672*** (0,0408)	0,2072*** (0,0412)	0,1618*** (0,0411)	0,2016*** (0,0414)
Vendedores	0,1679*** (0,0379)	0,2794*** (0,0445)	0,1603*** (0,0384)	0,2713*** (0,0445)
Agrícola	-0,2135*** (0,0396)	-0,0516 (0,0775)	-0,2058*** (0,0398)	-0,0447 (0,0789)
Produção e manutenção	0,1582*** (0,0243)	0,2195*** (0,0308)	0,1616*** (0,0246)	0,2195*** (0,0307)
Ocupação Informal		-0,0953*** (0,0264)		-0,0835*** (0,0264)
Ramos da atividade (Base: Serviços)				
Agrícola		-0,1749** (0,0759)		-0,1764** (0,0773)
Indústria de transformação e outras		-0,0362 (0,0338)		-0,0419 (0,0341)
Construção civil		-0,1248*** (0,0351)		-0,1130*** (0,0349)
Comércio e reparação		-0,1593*** (0,0335)		-0,1587*** (0,0334)
Outras atividades		0,0287 (0,0367)		0,0250 (0,0367)
Fatores locacionais				
Urbana	0,0985** (0,0389)	0,0923** (0,0388)	0,1008*** (0,0387)	0,0945** (0,0386)
Região metropolitana	-0,0236 (0,0187)	-0,0190 (0,0187)	-0,0228 (0,0187)	-0,0187 (0,0187)
Norte	-0,1569***	-0,1467***	-0,1550***	-0,1460***
Nordeste	-0,4062***	-0,3973***	-0,4096***	-0,4016***
Sul	-0,0053	-0,0058	-0,0035	-0,0036
Centro-Oeste	-0,0085	-0,0013	-0,0118	-0,0051
lambda	-0,4258***	-0,4138***	-0,4041***	-0,3954***
Constante	2,1590***	2,1907***	2,1830***	2,2138***
Observações (N)	6.979	6.979	6.979	6.979
N expandido (em milhões)	24,56	24,56	24,56	24,56
R2 ajustado	0,49	0,50	0,49	0,49

Fonte: Elaboração própria a partir dos microdados da PNAD/PETab 2008.

Legenda: *p<0,10; **p<0,05; ***p<0,01.

Tabela C.4 – Análise de Robustez – diferentes especificações de instrumentos para os modelos de penalização e diferenciação compensatória

Detalhamento	MQO	Variáveis Instrumentais (VI)						
		1	2	3	4	5		
Painel A: penalização								
Coeficiente (Fumante)	-0,1053	-0,2211	-0,2312	-0,3001	-0,3016	-0,2479		
Erro-padrão	0,0212	0,0599	0,0598	0,0851	0,0850	0,0526		
R2-ajustado	0,4918	0,4885	0,4879	0,4826	0,4824	0,4869		
Resistência dos instrumentos								
<i>Ajustamento do primeiro estágio</i>								
R2-ajustado		0,2477	0,2476	0,1609	0,1609	0,2838		
R2-parcial		0,1685	0,1685	0,0726	0,0727	0,2088		
Estatística-F robusta		190,21	126,81	166,42	111,32	111,08		
P-valor(F)		0,0000	0,0000	0,0000	0,0000	0,0000		
Instrumentos								
Proibido fumar no domicílio				X	X	X		
<i>Crença/percepção</i>								
Gera doenças aos outros				X	X	X		
Gera doenças para si					X	X		
Fumantes na família		X	X			X		
Ex-fumantes na família		X	X			X		
Preço do cigarro			X			X		
Atributos individuais								
	X	X	X	X	X	X		
Fatores locacionais								
	X	X	X	X	X	X		
Controle: ex-fumantes								
	X	X	X	X	X	X		
Número de observações	6.979	6.979	6.979	6.979	6.979	6.979		
		Não-Fumante			Fumante			
		VI				VI		
	MQO	1	2	3	MQO	1	2	3
Painel B: diferenciação								
Coeficiente (Índice de risco)	0,0491	0,1523	0,1552	0,1357	0,0335	0,1045	0,1016	0,0813
Erro-padrão	0,0040	0,0123	0,0123	0,0105	0,0066	0,0249	0,0248	0,0224
R2-ajustado	0,4584	0,4362	0,4350	0,4428	0,4212	0,4103	0,4111	0,4163
Resistência dos instrumentos								
<i>Ajustamento do primeiro estágio</i>								
R2-ajustado		0,1303	0,1307	0,1638		0,0911	0,0907	0,1044
R2-parcial		0,0950	0,0955	0,1300		0,0647	0,0649	0,0790
Estatística-F robusta		1529,91	574,82	583,08		466,59	176,07	182,60
P-valor(F)		0,0000	0,0000	0,0000		0,0000	0,0000	0,0000
Instrumentos								
TOA		X	X	X		X	X	X
TOA ²				X				X
Casal sem filhos		X	X	X		X	X	X
Casal com filhos		X	X	X		X	X	X
Chefe do domicílio			X	X			X	X
Renda de não-trabalho			X	X			X	X
Esposa com dificuldades físicas			X	X			X	X
Filho com dificuldades físicas			X	X			X	X
Atributos individuais								
	X	X	X	X	X	X	X	X
Fatores locacionais								
	X	X	X	X	X	X	X	X
Controle: ex-fumantes								
	X	X	X	X	X	X	X	X
Número de observações	22.088	22.088	22.088	22.088	8.153	8.153	8.153	8.153

Fonte: Elaboração própria a partir dos microdados da PNAD/PETab 2008 e do AEAT/MPS 2008.

Tabela C.3 – Definição das variáveis usadas na pesquisa por tipo de modelo

Variáveis		Definição das variáveis	M(1)	Modelos	
			λ	Z1	Z2
Indicadores ocupacionais					
Salário-hora	Participação na força de trabalho	Renda do trabalho principal na semana de referência por horas trabalhadas (em logaritmo)	X	X	X
Sindicalizado		1 se o indivíduo possui rendimentos positivos e não nulos, 0 caso contrário		X	
IRAT		1 se o trabalhador é sindicalizado, 0 caso contrário	X		X
<i>Grupos ocupacionais</i>					
Dirigentes		Índice de risco de acidentes do trabalho na divisão econômica por 100 trabalhadores		X	X
Ciências e Artes		1 se dirigentes em geral no trabalho principal, 0 caso contrário (c.c.)	X	X	X
Técnicos		1 se profissionais das ciências e das artes, 0 c.c.	X	X	X
Serviços administrativos		1 se técnicos de nível médio, 0 c.c.	X	X	X
Serviços*		1 se trabalhadores de serviços administrativos, 0 c.c.	X	X	X
Vendedores		1 se trabalhadores dos serviços, 0 c.c.	X	X	X
Agrícola		1 se vendedores e prestadores de serviço do comércio, 0 c.c.	X	X	X
Produção e manutenção		1 se trabalhadores agrícolas, 0 c.c.	X	X	X
Estabilidade (<i>tenure</i>)		1 se trabalhadores da produção de bens e serviços e de reparação e manutenção, 0 c.c.	X	X	X
Atributos pessoais					
Faixa de escolaridade		Tempo de serviço em anos no emprego atual	X		
<EF		1 se menos que ensino fundamental completo, 0 c.c.	X	X	X
<EM		1 se ensino fundamental completo ou médio incompleto, 0 c.c.	X	X	X
<ES		1 se ensino médio completo ou superior incompleto, 0 c.c.	X	X	X
ES*		1 se ensino superior completo ou mais anos de estudo, 0 c.c.	X	X	X
Raça		1 se o trabalhador é da cor branca, 0 caso contrário	X	X	X
Idade		Idade em anos	X	X	X
Fumante		1 se o trabalhador é fumante, 0 caso contrário	X	X	X
Ex-fumante		1 se o trabalhador já fez uso de produto de tabaco; 0 caso contrário			
Tamanho da família		Componentes da família (exclusive pensionista, empregado doméstico ou parente do empregado doméstico)	X		
Dificuldades físicas		1 se tem dificuldade, por problema de saúde, para levantar objetos, praticar esportes ou realizar trabalhos pesados; 0 c.c.	X		
<i>Rendimento de outras fontes</i>					
Pensão		Pensão de instituto de previdência ou do governo federal, ou outro tipo (em logaritmo)	X		
Aluguel		Rendimento de aluguel (em logaritmo)	X		
Outros		Juros, dividendos, programas sociais ou outros rendimentos, abono ou doação (em logaritmo)	X		
Fatores locais					
Urbana		1 se o indivíduo reside em área urbana, 0 c.c.	X	X	X
Região metropolitana		1 se o indivíduo reside em região metropolitana, 0 c.c.	X	X	X

Tabela C.3 – Definição das variáveis usadas na pesquisa por tipo de modelo (continuação)

Variáveis	Definição das variáveis	Modelos			
		M(1)	λ	Z1	M(2) Z2
<i>Macroregião</i>					
Norte	1 se o indivíduo reside na região Norte, 0 c.c.	X	X		X
Nordeste	1 se o indivíduo reside na região Nordeste, 0 c.c.	X	X		X
Sudeste	1 se o indivíduo reside na região Sudeste, 0 c.c.	X	X		X
Sul	1 se o indivíduo reside na região Sul, 0 c.c.	X	X		X
Centro-Oeste	1 se o indivíduo reside na região Centro-Oeste, 0 c.c.	X	X		X
Instrumentos: penalização					
Proibido fumar no domicílio	1 se é proibido fumar em qualquer lugar do domicílio, 0 c.c.			X	
<i>Crença/percepção</i>					
Gera doenças aos outros	1 se o indivíduo acredita que fumar gera doença para terceiros, 0 c.c.			X	
Gera doenças para si	1 se o indivíduo acredita que fumar gera doenças graves para si, 0 c.c.			X	
Fumantes na família	Quantidade de pessoas fumantes que residem com o i-ésimo indivíduo (exclui a própria pessoa)			X	
Ex-fumantes na família	1 se existe algum ex-fumante no domicílio (exclui a própria pessoa), 0 c.c.			X	
Preço do cigarro	Média do preço do pacote (maço) do cigarro no estado			X	
Instrumentos: diferenciação					
Chefe do domicílio	1 se o indivíduo é o morador de referência do domicílio, 0 c.c.		X		X
<i>Tipo de família</i>					
Casado sem filhos	1 se a pessoa é casada mas sem filhos, 0 c.c.		X		X
Casado com filhos	1 se a pessoa é casada mas com filhos, 0 c.c.		X		
Renda total do não-trabalho	Rendimento total de outras fontes não ligadas ao trabalho atual (em log)				X
TOA	Total de ocupados em todas as outras atividades na mesma divisão econômica, exceto a atividade indivíduo i				X
Cônjuge: dificuldades físicas	1 se o cônjuge possui dificuldades físicas (tal como descrito acima); 0 c.c.				X
Filho: dificuldades físicas	1 se o filho do indivíduo tem dificuldades físicas (tal como descrito acima); 0 c.c.				X

Fonte: Elaboração própria.

Coluna **M(1)** representa as variáveis utilizadas para as equações de rendimento, com ênfase na obtenção da penalização salarial promovida pelo o uso do cigarro.

Coluna λ indica as variáveis usadas para a correção do viés de seleção amostral.

Coluna **Z1** reporta a descrição das variáveis instrumentais adotadas para os modelos MQ2E e RQVI em M(1).

Coluna **M(2)** mostra as variáveis utilizadas para a análise da diferenciação compensatória de atividades de risco entre fumantes e não-fumantes.

Coluna **Z2** reporta a descrição das variáveis instrumentais adotadas para os modelos MQ2E e RQVI em M(2).